

Dissertation zur Erlangung des Doktorgrades  
der Fakultät für Chemie und Pharmazie  
der Ludwig-Maximilians-Universität München

**Investigations of Click Chemistry Approaches towards Diagnostic  
and Therapeutic Applications of Nucleic Acids**

**Eva Sophie Schönegger**

aus  
Klagenfurt, Österreich

2024

## Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Thomas Carell betreut.

## Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 07.03.2024

.....

Eva Sophie Schönegger

Dissertation eingereicht am: 07.03.2024

1. Gutachter: Prof. Dr. Thomas Carell

2. Gutachter: Dr. Pavel Kielkowski

Mündliche Prüfung am: 12.04.2024

# Table of Contents

Publications List.....	I
Acknowledgements.....	III
Abbreviations.....	V
Abstract.....	VI
1. Introduction.....	1
1.1 Click Chemistry.....	1
1.1.1 Huisgen Azide Alkyne Cycloaddition.....	2
1.1.2 Copper Catalyzed Azide Alkyne Cycloaddition (CuAAC).....	3
1.1.3 Ruthenium Catalyzed Azide-Alkyne Cycloaddition (RuAAC).....	6
1.1.4 Strain Promoted Azide-Alkyne Cycloaddition (SPAAC).....	7
1.1.5 Accelerated Metal Free Click Reaction: Tetrazine Ligation.....	8
1.2 Click Chemistry on Nucleic Acids.....	10
1.3 Application of Click-Functionalized Nucleic Acids.....	15
1.3.1 Therapeutic Applications.....	15
1.3.2 Diagnostic Applications.....	17
1.3.2.1 Metabolic Labelling.....	17
1.3.2.2 Next and Third Generation Sequencing.....	18
2. Aim of this Thesis.....	19
3. Published Works.....	20
3.1. First Publication.....	20
3.2. Second Publication.....	29
4. Unpublished Work.....	38
4.1. Click Chemistry Based Library Preparation for Direct cDNA Sequencing.....	38
4.1.1. Prologue.....	38
4.1.2. Results and Discussion.....	39
4.2. Unpublished Data towards the 2 <sup>nd</sup> Publication.....	42
4.2.1. Prologue.....	42
4.2.2. Results and Discussion.....	43
4.3 Click Modified Cu(II)-Dependent Nucleic Acid-Based Artificial Ribonuclease for Targeted Cleavage of SARS-CoV-2 Genomic RNA.....	54
4.3.1 Prologue.....	54
4.3.2 Results and Discussion.....	57

4.4	Investigations towards Copper-Free Click Chemistry for Optimized mRNA Vaccine Development.....	70
4.4.1	Quantification of 3'-N <sub>3</sub> Labelling and SPAAC with DBCO-Trimannose-Moiety for baseclick's mRNA Vaccine Development .....	70
4.4.1.1	Prologue.....	70
4.4.1.2	Results and Discussion.....	72
4.4.2	Circularization of an Oligoribonucleotide <i>via</i> SPAAC for Potential Circular mRNA Application.....	78
4.4.2.1	Prologue.....	78
4.4.2.2	Results and Discussion.....	79
5.	Thesis Discussion.....	85
6.	Experimental Section of Unpublished Work.....	90
6.1.	General Experimental Section and Materials.....	90
6.2.	Click chemistry-based library preparation and direct cDNA sequencing.....	93
6.3	Experimental section of unpublished data towards the 2 <sup>nd</sup> Publication .....	95
6.3.1	Chemical Synthesis .....	95
6.3.1.1	Synthesis of 3-(2-(2-(2-tosyloxyethoxy)ethoxy)ethoxy)-N-[3-oxo-N-(DBCO)propyl] propanamid ( <b>24</b> ).....	96
6.3.1.2	Synthesis of $\gamma$ -DBCO-PEG4-ATP ( <b>25</b> ).....	97
6.3.1.3	Synthesis of $\gamma$ -allyl-ATP ( <b>26</b> ) .....	98
6.3.1.4	Synthesis of 3'-azido-2',3'-dideoxyadenosine monophosphate ( <b>27</b> ) .....	99
6.3.2	T4 PNK Catalyzed 5'-End Labelling Reaction .....	100
6.4	Click-Modified Cu(II)-Dependent Nucleic Acid-Based Artificial Ribonuclease for Targeted Cleavage of SARS-CoV-2 Genomic RNA .....	102
6.4.1	Synthesis of Internal C8-Alkyne Modified Oligonucleotides.....	102
6.4.2	CuAAC of N <sub>3</sub> -Cu-Ligand to Alkyne Modified Oligonucleotides.....	104
6.4.3	Binding Affinity Analysis of C8-Alkyne and Cu-Binding Ligand Clicked Oligonucleotides towards Target, Off-Target Sequences <i>via</i> nPAGE.....	111
6.4.3.1	Binding Affinity Analysis of C8-Alkyne and Cu- Ligand Clicked Oligonucleotides towards Target, Off-Target without Cu(II) -Source .....	112
6.4.3.2	Binding Affinity Analysis of Selected Cu-Ligand-Clicked Oligonucleotides towards Target, Off-Target with Cu(II) Source .....	112
6.4.4	Cleavage Analysis of Cu-Ligand-Clicked Oligonucleotides towards Target, Off-Target.....	113
6.4.5	Binding Affinity Measurement and Determination of Dissociation Constant K <sub>D</sub> <i>via</i> Microscale Thermophoresis (MST) .....	114

6.5 Investigations towards Copper-Free Click Chemistry for Optimized mRNA Vaccine Development .....	116
6.5.1 Quantification of 3'-N <sub>3</sub> Labelling and SPAAC with DBCO-Trimannose-Moiety for baseclick's mRNA vaccine Development.....	116
6.5.2 Circularization of an RNA-RNA Oligoribonucleotide Assembly <i>via</i> SPAAC and Enzymatic Ligation for Potential Circular mRNA Application.....	122
7. References .....	124
8. Appendices.....	133
8.1 Supporting Information of Published Work .....	133
8.2 NMR Spectra of Unpublished Work .....	183

## Publications List

### Publications submitted towards this cumulative dissertation:

1. E. S. Schönegger, A. Crisp, M. Müller, J. Fertl, S. Serdjukow, S. Croce, M. Kollaschinski, T. Carell, T. Frischmuth, *Bioconjugate Chemistry* **2022**, *33*, 1789-1795.
2. E. S. Schönegger, A. Crisp, M. Radukic, J. Burmester, T. Frischmuth, T. Carell, *ChemBioChem* **2023**, e202300701.

### Conference Presentations:

Poster presentation: “*Modified mRNA SARS-CoV-2 Vaccine Development for Targeted Delivery*”. Third Annual RNA Therapeutics: From Concept to Clinic (June, 2021). Virtual Online Meeting.

Oral presentation: “*Click Chemistry applied for NGS and mRNA therapeutics*”. NATURE-ETN 2<sup>nd</sup> Meeting (September, 2021). Virtual Symposium.

Poster presentation: “*Click Chemistry enables rapid amplification of full-length reverse transcripts for long read Next Generation Sequencing*”. SCNAC 2022 – Symposium on Chemistry of Nucleic Acid Components (June, 2022). Český Krumlov, Czech Republic.

Oral presentation: “*Oligonucleotide based Therapeutic against SARS-CoV-2*”. NATURE ETN 4<sup>th</sup> Meeting (October, 2022). Dublin City University, Ireland.

Poster presentation: “*Click Chemistry enables rapid amplification of full-length reverse transcripts for long read Next Generation Sequencing*”. The MINERVA Center for Bio-hybrid Complex Systems scientific school and conference on “Bioinspired Complex Systems from Basic Science to Practical Applications” (November, 2022). Israel.

Poster presentation: *“Orthogonal End Labelling of Oligonucleotides through Dual Incorporation of Click-Reactive NTP Analogues”*. NATURE ETN 6<sup>th</sup> Meeting (November, 2023). Chimie ParisTech, France.

## Acknowledgements

During the course of completing this doctoral dissertation, a number of individuals have consistently offered me their support and guidance, whom I would like express my sincere gratitude.

First and foremost, I would like to express my profound appreciation to Dr. Thomas Frischmuth for his support and granting me the opportunity to pursue my doctoral studies at the baseclick GmbH, and becoming part of the NATURE ETN project, that has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 861381.

I also want to convey my sincere gratitude to my University supervisor Prof. Dr. Thomas Carell for welcoming me as PhD student and offering me a place within this group at the LMU. I deeply appreciate his helpful advices, guidance and motivation throughout my research endeavors.

I am profoundly indebted to Dr. Antony Crisp as his vital contributions played a pivotal role in the success of my doctoral work. His tremendously valuable advices, constant motivation, expertise and guidance extending beyond the Bavarian border, have provided indispensable support. I am deeply grateful for having Antony by my side and appreciated his scientific curiosity, interest and his help in all aspects of my doctoral studies.

Moreover, I would like to sincerely thank Dr. Markus Müller for his support during my doctoral studies. His valuable feedbacks and fruitful discussions helped me in the advancement of my projects. I am tremendously grateful for his support.

My colleagues from the baseclick GmbH I wish to thank for their support and helpful advices, especially Dr. Fabio Spada, Dr. Stefan Wiedemann and former colleague Dr. Stefano Croce. It was a pleasure to work together and learn from you. Moreover, would like to thank my students Jonas Burmester and Daniela Staudacher for their motivation and hard work which contributed to the progress of my research projects. Also, I want to thank Dr. Marco Radukic, from the Bielefeld University, for our fruitful and successful collaboration, together with Antony. Prof. Dr. Andrew Kellett and Dr. Brionna McGroman I would like to thank for welcoming me as a visiting student during my secondment at the DCU and introducing me to their scientific expertise. Moreover, I am very grateful for the support of Dr. Nada Raddouri, regarding all organizational matters.

One person that made my everyday lab-life more pleasant was Dr. Luis Montiel, my PhD partner (in crime :). I more than enjoyed working in the lab next to you, our “dirty-water” breaks, complaining together about the price increases at “Vor-Ort”, and (un)successful experiments and HPLC data storage issues, but most important I am grateful for our developed friendship that extended the baseclick lab-walls.

It was a great pleasure to be part of the Carell group, which undoubtedly enriched my time as PhD student, scientifically as well as personally. All colleagues were extraordinarily cooperative, which I deeply appreciated and I am especially happy to have met Kata and Giacomo, who always accepted my annoying HPLC booking messages, and joined me for after lab-dinner and/or boxing classes.

Additionally, I would like to express my gratitude to all ERS- colleagues from the NATURE ETN project. The moments we shared during our training weeks and conferences were truly unique and will be lasting memories of my PhD journey.

Deeply grateful I am for always having my friends from Austria by my side. You were and still are constants in my life which pushed me through ups and downs!

Last but not least I want to extend my heartfelt appreciation to my parents and my sister Änni. Your support and ongoing care (+ providing me with PhD-doping aka home-made products from Carinthia :) were deeply cherished.

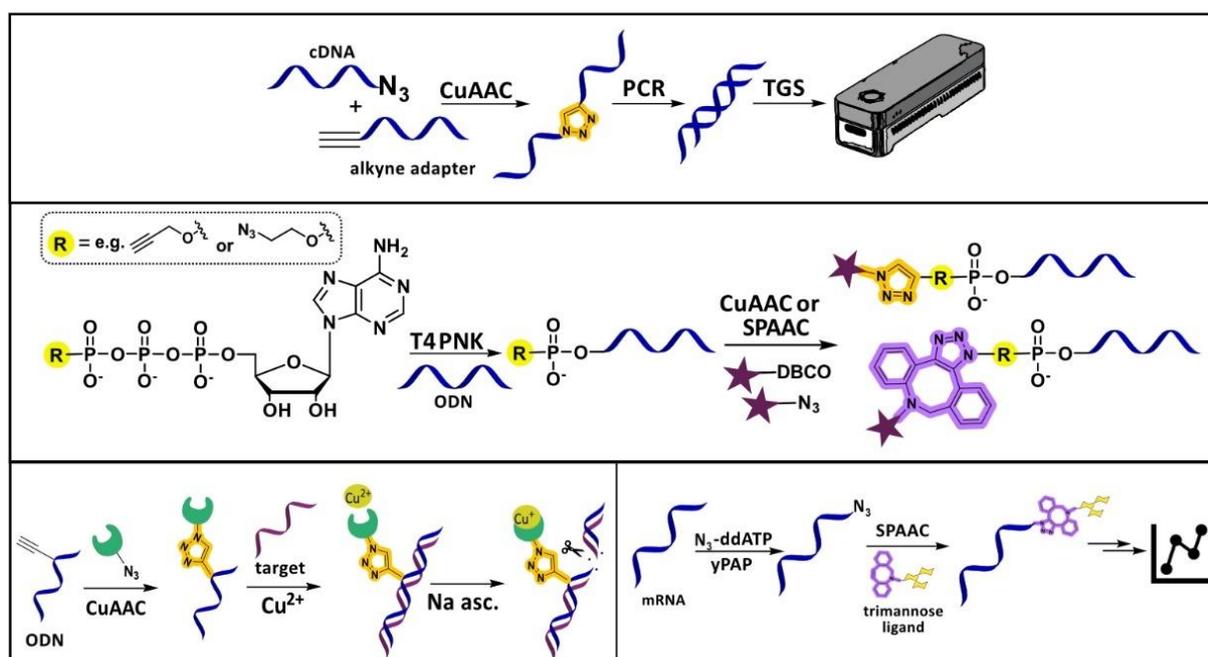
One significant acknowledgement is saved for the end- David, my biggest supporter during this PhD-journey. Your constant encouragements, advices, patience, support in all aspects and (very important :) suggests to chill now and then, were indispensable and truly valued.

## Abbreviations

A	adenine	equiv.	molar equivalent
abs.	absolute	G	guanine
AcOH	acetic acid	HBTU	2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
Ac <sub>2</sub> O	acetic anhydride	HSQC	heteronuclear single quantum coherence
<i>aq.</i>	aqueous	IBCF	isobutyl chloroformate
ADP	adenosine diphosphate	iBu	isobutyryl
ATP	adenosine triphosphate	MALDI TOF	matrix-assisted laser desorption/ionization time of flight
Bu <sub>4</sub> N <sup>+</sup>	tetrabutylammonium	M	molar [mol/l]
Bz	benzoate	Me	methyl
C	cytosine	MeOH	methanol
cDNA	complementary DNA	MHz	megahertz
CPG	controlled pore glass	<i>m/z</i>	mass to charge ratio
Cu(Cl <sub>4</sub> O) <sub>2</sub>	copper(ii) perchlorate	Na asc.	sodium ascorbate
CuSO <sub>4</sub>	copper sulfate	NaHCO <sub>3</sub>	sodium hydrogen carbonate
Cy3/5	cyanine 3/5	NMR	nuclear magnetic resonance
d	desoxy	rpm	revolutions per minute
DCA	dichloro acetic acid	<i>sat.</i>	saturated
DNA	deoxyribonucleic acid	T	thymidine
DBCO	dibenzocyclooctyne/dibenzoazacyclooctyne	THPTA	tris[(1-hydroxypropyl-1h-1,2,3-triazol-4-yl)methyl]amine
DCM	dichlormethane	TLC	thin layer chromatography
DIPEA	<i>n,n</i> -diisopropylethylamine	TOS-PEG4-acid	3-(2-(2-(2-(tosyloxy)ethoxy)ethoxy)ethoxy)propanoic acid
DMF	dimethylformamide	TRIS	tris(hydroxymethyl)aminomethane
DMSO	dimethyl sulfoxide	U	uracil
EDTA	ethylenediaminetetraacetic acid	%	percentage yield
ESI	electron spray ionisation	μ	mikro (10 <sup>-6</sup> )
EtOAc	ethyl acetate	δ	chemical shift
EtOH	ethanol	ε	molar extinction coefficient
Et <sub>3</sub> N	triethylamine	λ	wavelength

## Abstract

The work presented in this dissertation is aimed towards broadening the scope of click-functionalized nucleic acids for emerging diagnostic and therapeutic applications. Herein, investigations on the modification of nucleic acid building blocks through synthetic or chemoenzymatic approaches were conducted, encompassing nucleotides, oligonucleotides, and both DNA and RNA. Specifically, a novel click chemistry-based library preparation method for Third Generation Sequencing was successfully introduced. We developed a straightforward protocol, covering the full length of transcripts, as well as utilizing the well-known advantages of click chemistry and thus avoiding problematic template switching reactions. In a further exploration of post-synthetic nucleic acid click chemistry 5'-end labelling of nucleic acids with click-functionalized moieties, enabling subsequent click reactions of oligonucleotides with fluorescent dyes. This approach afforded the synthesis of  $\gamma$ -phosphate modified ATP analogues, the setup of a chemoenzymatic labelling assay and proof of applicability, which are presented within this work. Towards therapeutic application, a Cu(I) dependent click-modified artificial ribonuclease is introduced, aiming to enable a novel approach for target specific cleavage of the genomic RNA of SARS-CoV-2, thus inhibiting viral replication. Furthermore, investigations towards advancement of mRNA vaccine development, by copper-free click chemistry were conducted. Herein, the efficiency of a 2-step 3'end mRNA labelling and an approach to RNA circularization *via* SPAAC were examined.

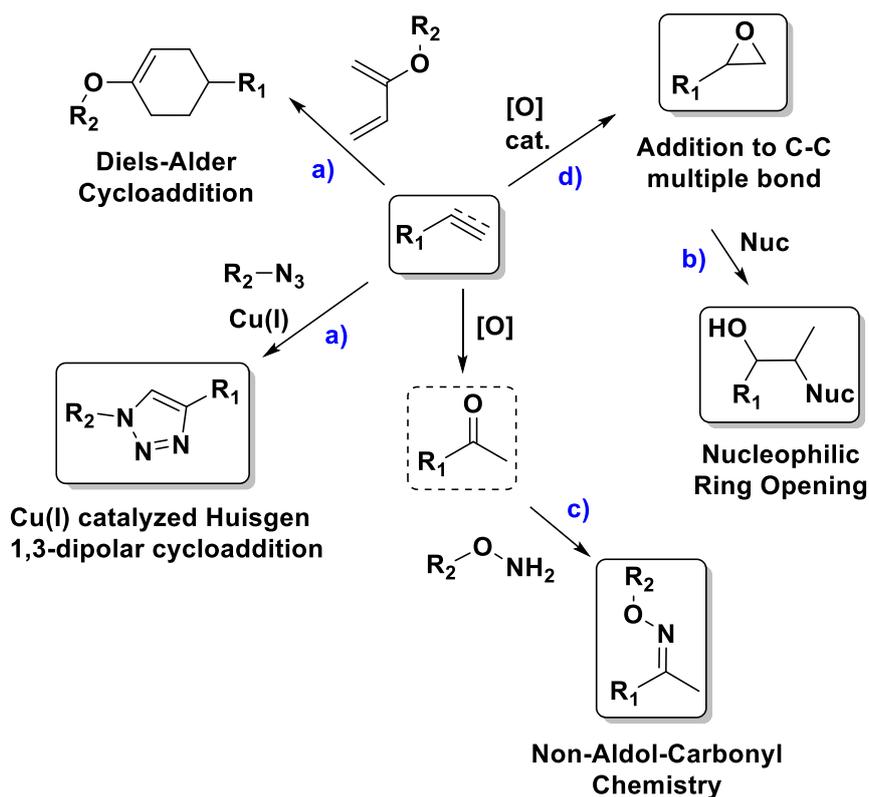


# 1. Introduction

## 1.1 Click Chemistry

The term “click chemistry” was coined by Barry Sharpless from the Scripps Research Institute, in 2001 and describes a class of chemical reactions, that fulfill specific criteria and are characterized by their high efficiency, selectivity, reliability and mild reaction conditions. <sup>[1]</sup> At the same time, independently from Sharpless, Morten Meldal from the University of Copenhagen, also discovered this groundbreaking reaction that links two molecules, an azide and an alkyne in a straightforward manner.<sup>[2]</sup> Since then, click chemistry has become an indispensable technique to efficiently join molecules together, even in living cells, without interfering with the cell function, which has been first shown by Carolyn Bertozzi from Stanford University.<sup>[3]</sup> For their discoveries towards “click chemistry”, which to this day, have had profound impact across the fields of chemistry, biology and related research areas, Bertozzi, Sharpless and Medal have been awarded for the Nobel Prize in Chemistry in 2022.<sup>[4]</sup>

As mentioned above, the concept of click chemistry was pioneered by Kolb, Finn and Sharpless, specifying a set of criteria. According to their definition, click reactions need to be modular, broad in scope, high-yielding, stereospecific and generate no or minor byproducts that can easily be removed without the need of chromatographic methods. Moreover, the term “click” conveys the notion of reaction conditions characterized by simplicity of execution, the use of readily accessible starting materials, compatibility with aqueous or non-aqueous solvents, that can easily be removed, and straightforward product isolation. The formed products further need to be stable under physiological conditions.<sup>[1]</sup> Sharpless and co-workers defined four groups of bond formation, that fulfill the criteria mentioned above: a) cycloadditions of unsaturated moieties (e.g. 1,3-dipolar cycloadditions, Diels-alder reactions), b) nucleophilic substitution, especially ring-opening reactions of strained heterocyclic electrophiles (e.g. epoxides, aziridines), c) non-aldol-type carbonyl chemistry (e.g. urea, thiourea, amide formation) d) additions to C-C multiple bonds (e.g. epoxidation, dihydroxylation, aziridination, Michael addition) (Figure 1)

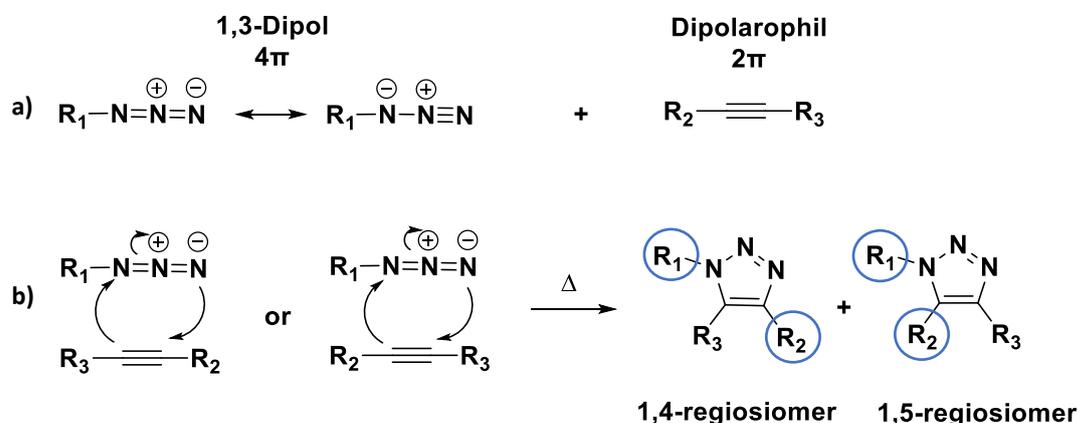


**Figure 1.** Schematic representation of reactions following the criteria of Click Chemistry. **a)** Cycloadditions of unsaturated moieties. **b)** Nucleophilic substitution **c)** Non-aldol-type carbonyl chemistry. **d)** Additions to C-C multiple bonds. The Figure was prepared in reference to Moses *et al.*<sup>[5]</sup>

### 1.1.1 Huisgen Azide Alkyne Cycloaddition

The Huisgen cycloaddition, a condensation reaction between organic azides with alkyne groups forming 1,2,3-triazole linkages, has been widely exploited by organic chemists for decades. The reaction is a concerted process, involving the pericyclic (2+3) cycloaddition of a dipolarophile (delivering 2π electrons) with a 1,3-dipole (reacting as 4π system), leading to the formation of a 5-membered (heterocyclic) ring (Figure 2).<sup>[6-8]</sup> The alkyne and azide moieties are chemically unreactive towards most functional groups of biological substrates, such as nucleic acids and proteins, and can be introduced into the scaffold of large molecules in a straightforward manner.<sup>[9]</sup> Nevertheless, the widespread utilization of the Huisgen cycloaddition reaction was hindered due to the requirement of high temperatures, extended reaction times, and the lack of regioselectivity, since both, 1,4- and 1,5-cycloadducts are formed (Figure 2), whereas the regioselectivity of the reaction depends on the electronic and steric effects of the reaction partners and is more or less predictable.<sup>[9-10]</sup> The slow conversion

of the reaction can be attributed to the inherent stability of canonical alkynes. Only electron-deficient alkynes can undergo this noncatalyzed cycloaddition through conjugate addition mechanisms.<sup>[9]</sup>

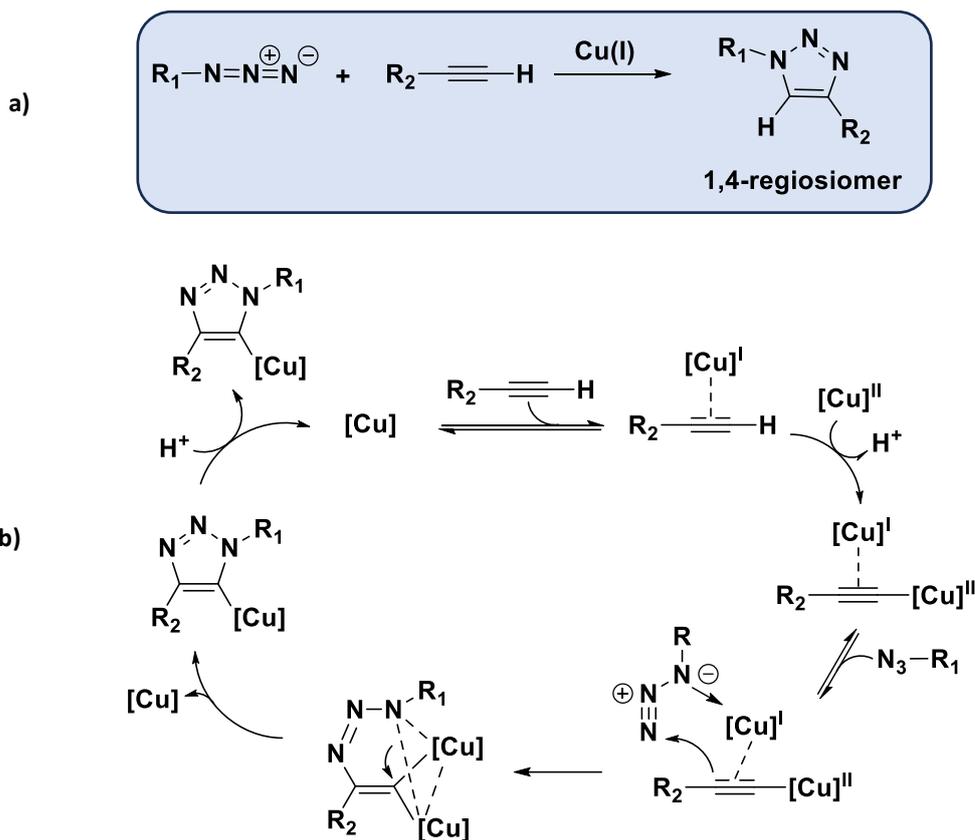


**Figure 2.** a) Starting materials for (3+2) cycloaddition. b) Mechanism of the Huisgen 1,3 dipolar cycloaddition.<sup>[8]</sup>

### 1.1.2 Copper Catalyzed Azide Alkyne Cycloaddition (CuAAC)

A significant breakthrough was achieved by Meldal and Sharpless as they simultaneously addressed the limitations of the Huisgen cycloaddition by introducing copper(I) as catalyst.<sup>[2, 10]</sup> Their independent studies, reported in 2002, showed that Cu(I) accelerates the kinetics of the azide-alkyne cycloaddition by  $10^7$ - $10^8$  fold and further leads to the formation of only one regioisomeric product, namely the 1,4-disubstituted-1,2,3 triazole (Figure 3a).<sup>[1, 11]</sup> The catalytic active Cu(I) source can be generated directly by the incorporation of Copper(I) salts, such as CuI or CuBr, the oxidation of Cu(0) precatalysts, e.g. elementary copper, or prepared in situ by reduction of Cu(II) salts, such as  $CuSO_4 \cdot 5H_2O$ , with sodium ascorbate.<sup>[9, 12]</sup> The latter one is more favorable, since the Cu(II)/ascorbate aqueous system is more reliable (producing purer Cu(I) species), cheaper and does not necessarily require co-solvents, which makes this system ideal for reactions with nucleic acids.<sup>[9-10]</sup>

Initially Sharpless *et al.*, proposed that the stepwise Cu(I) catalyzed azide-alkyne cycloaddition (CuAAC), is catalyzed by a mononuclear Cu(I) species.<sup>[10]</sup> In 2013, Fokin *et al.* however showed, based on DFT calculations that a dinuclear copper intermediate enables the CuAAC (Figure 3b).<sup>[13]</sup>

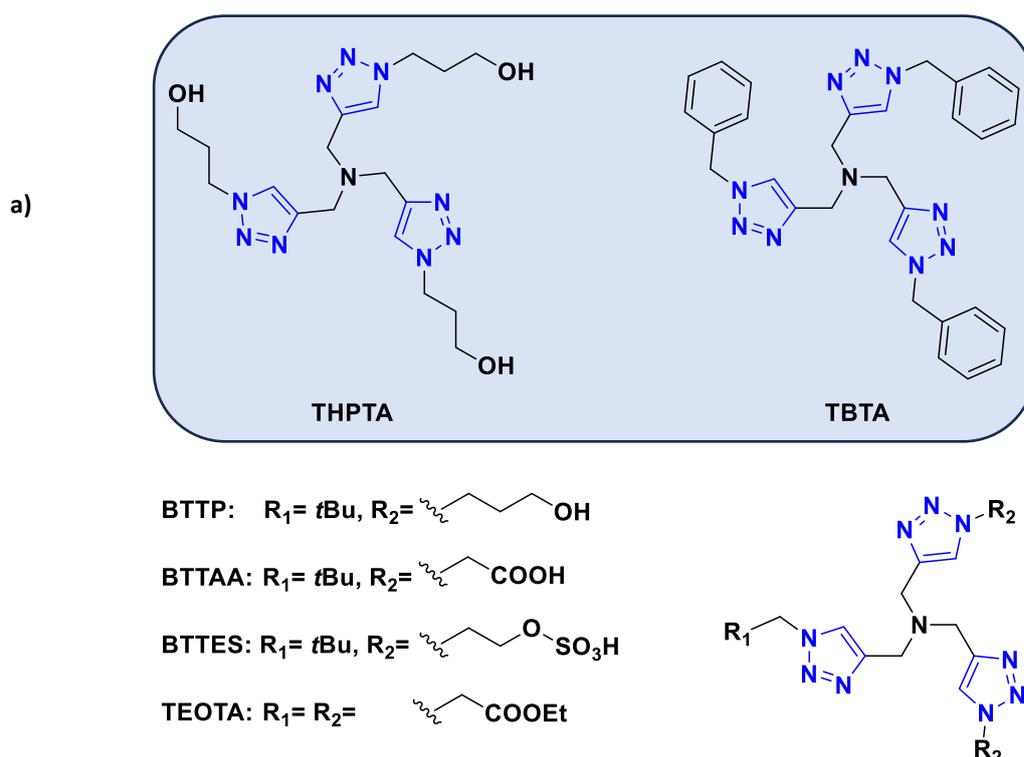


**Figure 3.** a) Schematic representation of the Cu(I) catalyzed azide alkyne cycloaddition (CuAAC) b) Proposed mechanism of the CuAAC with two copper atoms by Fokin *et al.*<sup>[13]</sup>

Following its initial discovery, the Cu(I)-catalyzed azide-alkyne cycloaddition has gained widespread use across the realms of chemistry, biology, biochemistry, and biotechnology, serving as a potent and versatile tool for multiple applications.<sup>[9]</sup> Since this chemical reaction does not interact and interfere with a biological system, involves functional groups that are not naturally present in biological systems and occurs at physiological conditions in a highly selective and highly yielding manner, the CuAAC fulfills the requirements of a bioorthogonal reaction.<sup>[14-15]</sup>

Since Cu(II) displays the most stable oxidation state of copper, however Cu(II) is catalytically inactive, efforts to further stabilize the Cu(I) species have been investigated, where the incorporation of polytriazole-based ligands showed, that the CuAAC can be significantly improved (up to several thousand times faster;  $10^0$ - $10^4$   $M^{-1} s^{-1}$ ).<sup>[9, 16-18]</sup> Moreover, by the use of these ligands, the load of copper catalyst can be lowered significantly, thus leading to decreased cytotoxicity.<sup>[19]</sup> Particularly tris(triazole)-amine based ligands are widely used, especially for CuAAC on nucleic acids, where the scaffold of the ligand is chosen depending on the solubility of the applied chemicals and corresponding solvent conditions.<sup>[9]</sup> For the use of

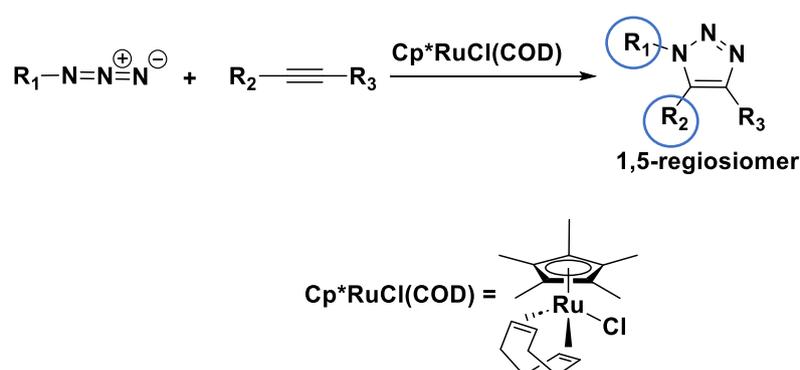
CuAAC in aqueous solution, e.g. with  $\text{CuSO}_4$ , THPTA (tris(3-hydroxypropyltriazolylmethyl)amine) and for CuAAC in organic solvents, e.g. with  $\text{CuBr}$ , TBTA (tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine) have become the most popular ligands especially for click reactions on nucleic acids (Figure 4).<sup>[9, 20]</sup> More recently, alternative nontoxic ligands, such as BTES (a sulfonic acid tris(triazolyl)amine derivative), BTAA and BTTP, have been explored for their ability to catalyze CuAAC bioconjugations, both *in vitro* and *in vivo*.<sup>[9, 20-21]</sup>



**Figure 4. a)** Structures of THPTA (tris(3-hydroxypropyltriazolylmethyl)amine) and TBTA (tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine) ligand. **b)** Structures of further tris(triazolylamine) ligands to accelerate CuAAC.<sup>[9]</sup> BTTP (3-[4-((bis[(1-tert-butyl-1H-1,2,3-triazol-4-yl)methyl]amino)methyl)-1H-1,2,3-triazol-1-yl]propanol), BTTAA (2-[4-((bis[(1-tert-butyl-1H-1,2,3-triazol-4-yl)methyl]amino)methyl)-1H-1,2,3-triazol-1-yl]acetic acid), BTES (2-[4-((bis[(1-tert-butyl-1H-1,2,3-triazol-4-yl)methyl]amino)methyl)-1H-1,2,3-triazol-1-yl]ethyl hydrogen sulfate), TEOTA (tris[(1-(2-ethoxy-2-oxoethyl)-1H-1,2,3-triazol-4-yl)methyl]amine).

### 1.1.3 Ruthenium Catalyzed Azide-Alkyne Cycloaddition (RuAAC)

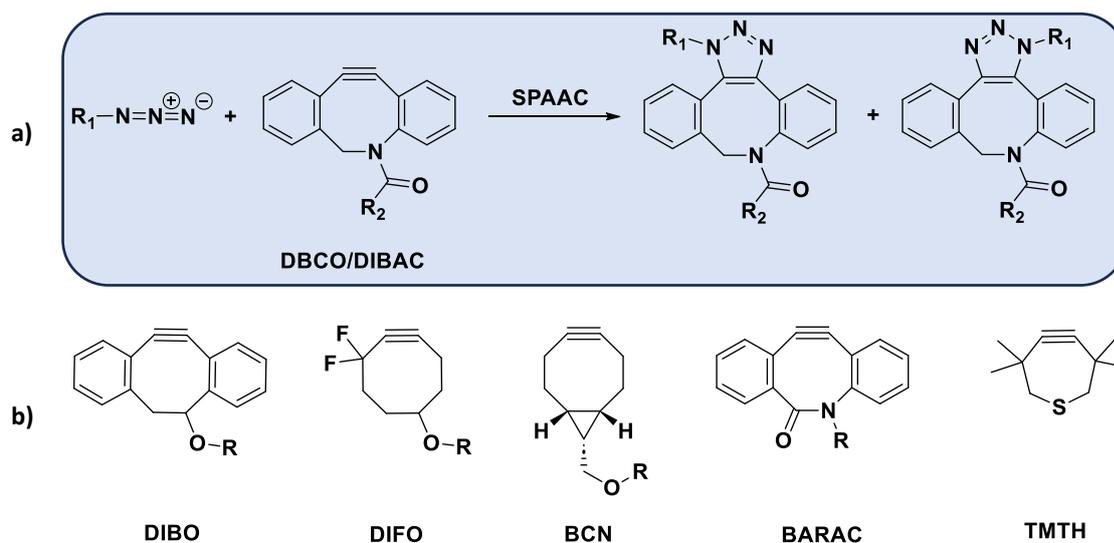
While by CuAAC only the 1,4-disubstituted 1,2,3 triazole is provided with terminal alkynes, the ruthenium catalyzed azide-alkyne cycloaddition (RuAAC) also enables the formation of 1,5-disubstituted 1,2,3-triazoles with internal alkynes ( $R_2$ ,  $R_3$ , Figure 5). The selectivity, however depends on the Ru-catalyst, since only sterically hindered catalysts, such as pentamethylcyclopentadienyl derivative ( $Cp^*$ ) like  $[Cp^*RuCl_2]_2$ , or  $Cp^*RuCl(COD)$  obtain the 1,5-disubstituted triazole only, while other derivatives, such as  $Ru(OAc)_2(PPh_3)_2$  favor the formation of the 1,4-disubstituted regioisomer.<sup>[22-23]</sup> The mechanism of these two metal catalyzed azide-alkyne cycloadditions differ from each other, and thus the RuAAC, in contrast to CuAAC, also tolerates internal alkynes, consequently enabling the formation of fully-substituted 1,2,3-triazoles (Figure 5).<sup>[22]</sup> Due to its more demanding reaction conditions, the utilization of RuAAC has been hindered compared to the CuAAC. Some investigations however require 1,5-disubstituted 1,2,3-triazoles, for example Tom Brown and co-workers studied the different 1,2,3-triazole regioisomers in term of their potential for mimicking phosphodiester backbones in nucleic acids.<sup>[9, 24]</sup> Their study showed that the 1,5-disubstituted 1,2,3-triazole, prepared by RuAAC, display higher replication kinetics than e.g. the 1,4-regioisomer and it is replicated precisely by polymerases.<sup>[24]</sup>



**Figure 5.** Schematic representation of the Ruthenium catalyzed azide-alkyne cycloaddition (RuAAC) with  $Cp^*RuCl(COD)$  catalyst ( $COD$ = cyclooctadiene) leading to the formation of the 1,5-disubstituted regioisomer.<sup>[22]</sup>

#### 1.1.4 Strain Promoted Azide-Alkyne Cycloaddition (SPAAC)

The metal-free, strain promoted azide-alkyne cycloaddition was first reported by Bertozzi and co-workers in 2004. With their approach they addressed the limitations of cytotoxic Cu-catalyzed azide-alkyne cycloadditions, to achieve *in vivo* applications and introduced a (3 + 2) cycloaddition between cyclooctynes and azides that can be performed under physiological conditions without any catalyst.<sup>[3]</sup> Initially the SPAAC was applied for ligations with glycans but over the years the bioconjugation has also been employed in various other applications such as for labelling or joining DNA strands.<sup>[9]</sup> The driving force for this reaction comes from significant bond angle deformation of the acetylene (almost 163°), and since this strain reduces the energy of the alkyne's LUMO (lowest unoccupied molecular orbital), it is more reactive and brings it closer in reactivity to the azide's HOMO (highest occupied molecular orbital).<sup>[3, 9]</sup> By enhancing the ring strain, e.g. by introducing electron-withdrawing groups (e.g. DIFO), or by attaching a 3-membered ring (BCN) the reactivity of cyclooctynes could be further improved. <sup>[9, 25-26]</sup> Most typically DBCO/DIBAC, DIBO, and BCN are employed in SPAAC, where their reaction rates are between 0.2–0.5 M<sup>-1</sup> s<sup>-1</sup>, depending on specific condition (Figure 6).<sup>[18, 25]</sup> There are BARAC and TMTH which are even more reactive (up to 4 M<sup>-1</sup> s<sup>-1</sup>), but exhibit poor stability, leading to ring destabilization and further unwanted side reactions with natural thiols.<sup>[26-27]</sup> Although the SPAAC has been widely applied in bioconjugations for diagnostic and therapeutic applications, the reaction exhibits some drawbacks, including much slower reaction rates compared to the CuAAC and the formation of two regioisomeric triazoles when e.g. DBCO is incorporated as the reactive cyclooctyne (Figure 6a).<sup>[9, 25]</sup>

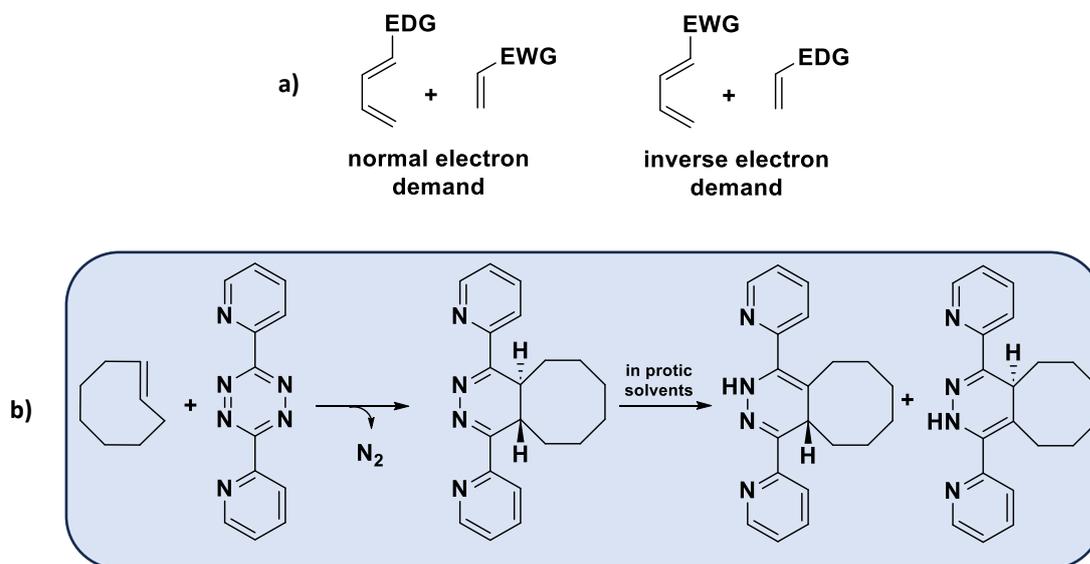


**Figure 6.** a) Schematic representation of the strain promoted azide-alkyne cycloaddition with DBCO/DIBAC (dibenzocyclooctyne/dibenzo-aza-cyclooctyne) leading to the formation of two regioisomers. b) Structures of typical strained alkynes employed in SPAAC. DIBO= dibenzocyclooctyne, DIFO= difluorocyclooctyne, BCN= bicyclo[6.1.0]non-4-yne, BARAC=biarylazacyclooctynones, TMTH= tetramethylthiacycloheptyne.

### 1.1.5 Accelerated Metal Free Click Reaction: Tetrazine Ligation

Another bioconjugation reaction, which does not require metal catalysts is the tetrazine ligation, based on inverse-electron-demand Diels–Alder reactivity. The origins of this reaction are based on the studies of Thalhammer and co-workers from 1990, where they described the kinetics of electron-deficient tetrazines with numerous electron rich dienophiles and demonstrated that their reactions with strained alkenes are exceptionally fast (Figure 7a).<sup>[28-29]</sup> Within these inverse-electron-demand Diels–Alder reactions, only  $N_2$  is produced as side product upon subsequent retro-[4 + 2] cycloaddition.<sup>[28-29]</sup> In 2008, Blackmann *et al.*, reported the efficient reactivity of the strained molecule trans-cyclooctene (TCO) with electron-deficient tetrazines in aqueous solutions and in the presence of model proteins (Figure 7b).<sup>[28]</sup> Due to its remarkable kinetic profile, with reaction rate constants ( $k$ ) spanning from  $10^3$  to  $10^6$   $M^{-1}s^{-1}$ , the TCO-tetrazine ligation, also known as third generation of click chemistry, has very quickly found numerous applications.<sup>[30]</sup> Further improvements towards the alkene-handles for tetrazine IEDDA labelling approaches have been made, where especially the small methylcyclopropene tag showed high selectivity and relative fast kinetics with tetrazines.<sup>[31-</sup>

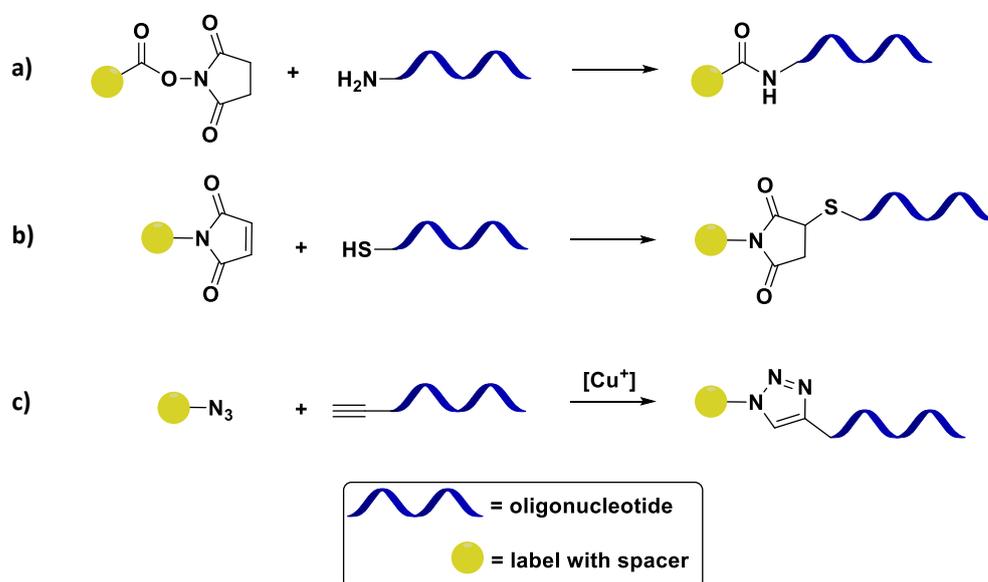
<sup>32]</sup> Even simpler, unstrained functionalities such as allyl-groups were able to react with modified tetrazines, in a lower conversion yield though.<sup>[32-33]</sup>



**Figure 7.** a) Comparison of normal and inverse electron demand Diels-Alder reaction. EWG= electron withdrawing, EDG= electron donating group.<sup>[34]</sup> b) Scheme of the tetrazine ligation (2-pyridyl-tetrazine) with trans cyclooctyne, introduced by Blackmann *et al.* <sup>[28]</sup>

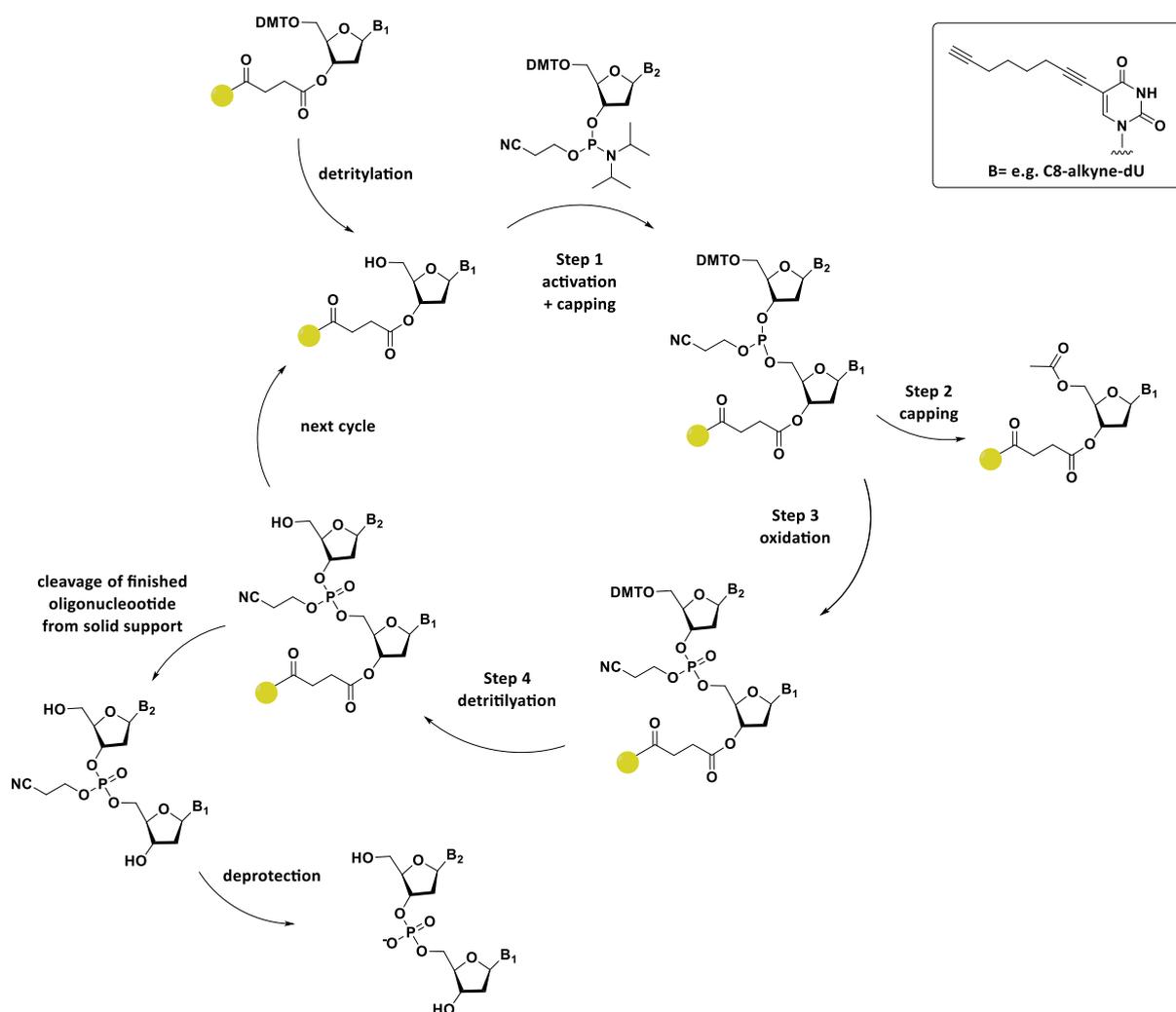
## 1.2 Click Chemistry on Nucleic Acids

Before azide-alkyne cycloadditions were applied on nucleic acids, the most common method to label oligonucleotides post-synthetically was achieved by the formation of amide bonds from amino-modified oligonucleotides (ODNs) and active ester, such as N-hydroxy succinimide ester (Figure 8a). However, these conjugations exhibit some drawbacks, including lower efficiency compared to CuAAC and sensitivity to ester hydrolysis in alkaline buffer ( $\text{pH} > 8$ ), which are used for the reaction.<sup>[9, 35]</sup> Another possibility to chemically modify nucleic acids is displayed by the Michael-type reaction between maleimide and thiol modified oligonucleotides (Figure 8b). The maleimide thiol reaction is however also limited by the electrophile instability ( $\text{C}=\text{C}$  bond of maleimide is highly electrophilic) and sensitivity of thiols towards dimerization.<sup>[9, 36-37]</sup> Due to the high chemical stability and orthogonality of organic alkynes and azides, the efficient alkyne-azide cycloaddition became of high interest for nucleic acid applications. Nowadays, the synthesis and incorporation of click-functionalized nucleic acid scaffolds are well-established and subsequent CuAAC reactions are usually high yielding and can be performed under mild reaction conditions (Figure 8c).<sup>[9, 35, 38-39]</sup>



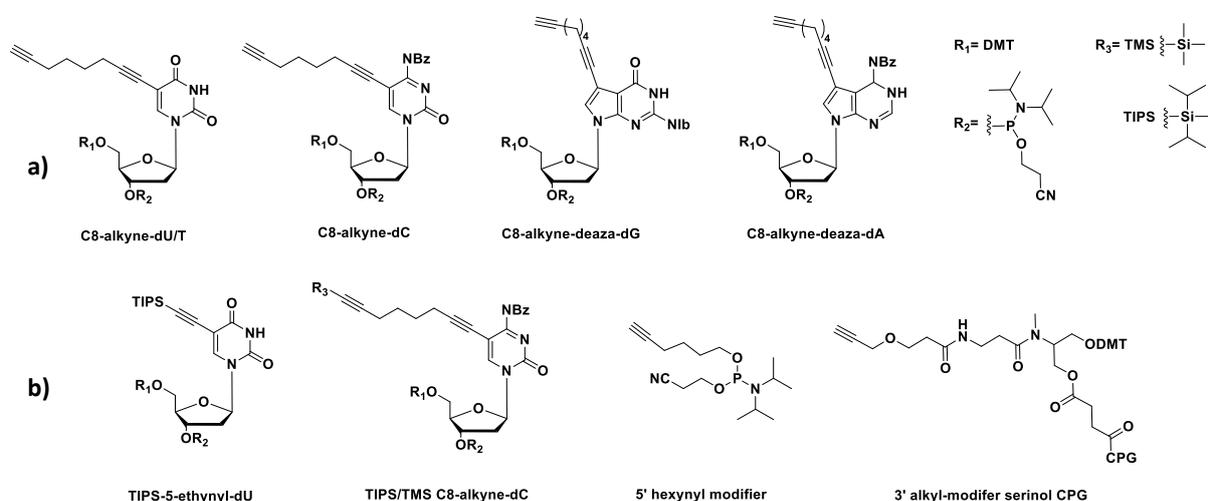
**Figure 8.** **a)** Amide linkage from a NHS ester and amino-functionalized oligonucleotide. **b)** Thiol-succinimide linkage from a maleimide and thiol-functionalized oligonucleotide. **c)** 1,4-disubstituted-1,2,3-triazole linkage from an azide and alkyne modified oligonucleotide.<sup>[35]</sup> The modifications on the oligonucleotides can be incorporated internally, at the 5' end and at 3' end respectively. Further maleimide modifications and NHS ester modifications can be incorporated into oligonucleotides and linked to thiol and amine probes respectively.<sup>[40]</sup>

The use of click-functionalized phosphoramidites and nucleoside triphosphates, for solid phase or enzymatic nucleic acid synthesis, respectively, reported by Carell *et al.*, and Seela *et al.*, in 2006, lead to a rapid development of CuAAC on nucleic acid.<sup>[38, 41]</sup> Oligonucleotides bearing a single alkyne modified base, can be prepared by incorporating C8-alkyne-modified phosphoramidites during solid phase synthesis (SPS, Figure 9, Figure 10a), where it has been shown that the 5-position of purine and the 7-position of 7-deazapurine nucleosides are the ideal positions for clickable functionalities, since these sites lie in the major groove of the DNA, offering steric freedom without altering base-pairing patterns.<sup>[9]</sup>



**Figure 9.** Schematic representation of the phosphoramidite oligonucleotide solid phase synthesis cycle.<sup>[42]</sup>

The C8-alkyne-dU-CE, with bearing a octadiynyl side chain with two triple bonds (further stabilizing DNA duplexes) and the 5-ethynyl-dU-CE (dU<sup>e</sup>-CE) phosphoramidites are due to their facile synthesis, the most applied alkyne modified building blocks for SPS.<sup>[41, 43]</sup> The 5-ethynyl-group, however, affords silyl protection during SPS, to avoid partial hydration and conversion to acetyl group formation during the deprotection cycle (Figure 10b). Further the reactivity of the click reaction on clustered ethynyl monomers is lower (~36%), compared to C8-alkyne linkers, due to steric shielding of the alkyne by the DNA backbone.<sup>[9, 38, 44-45]</sup> Classic approaches for CuAAC on nucleic acids use alkyne modified ODNs and azide probes, due to the reactivity of azide-functionalities with the P(III) groups (Staudinger reaction) during SPS.<sup>[9]</sup>

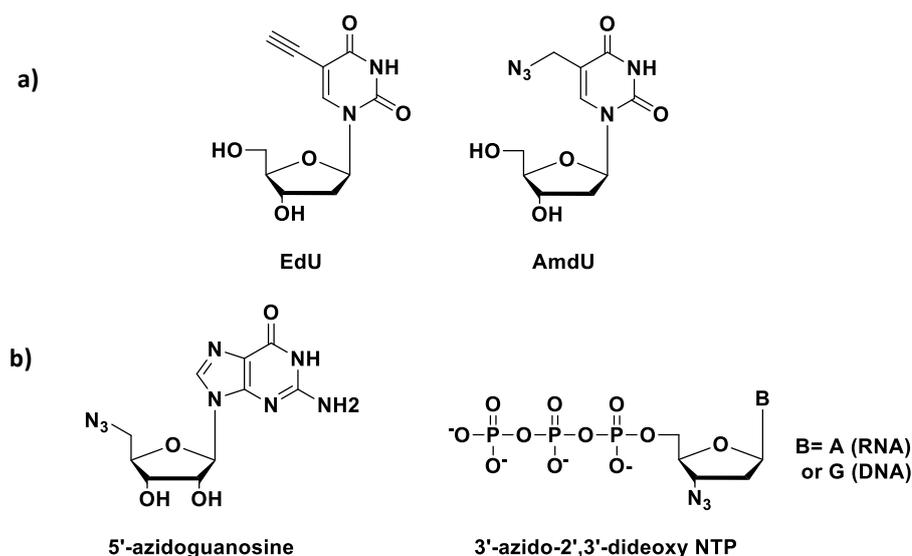


**Figure 10.** a) C8-alkyne modified phosphoramidites. b) TIPS-protected EdU-phosphoramidite, TIPS/TMS protected phosphoramidites for multiple labelling approach.<sup>[46]</sup> Alkyne modifiers for 5' and 3' end modifications.<sup>[9]</sup>

Multiple, sequential labelling with up to three different marker azides can be achieved by the incorporation of C8-alkyne modified, triisopropylsilyl (TIPS)-protected C8-alkyne modified and trimethylsilyl (TMS)-protected C8-alkyne modified phosphoramidites respectively, into oligonucleotides.<sup>[46]</sup> This approach was reported by Carell and co-workers, where the first click reaction is performed directly on the solid support. The resulting mono-modified oligonucleotide is then cleaved from the solid support by simultaneous cleavage of the TMS group (conc. NH<sub>3</sub> in H<sub>2</sub>O/EtOH). Next, after RP-HPLC purification, the 2<sup>nd</sup> click reaction is performed in solution and after precipitation of the doubly modified oligonucleotide, the TIPS group is cleavage of with TBAF (tetrabutylammonium fluoride). The 3<sup>rd</sup> click reaction in solution yields the triply labelled oligonucleotides. With the incorporation of non-nucleosidic alkyne phosphoramidite monomers, alkyne modifications can be introduced on the 5' or 3' -

end of the oligonucleotides, e.g. 5'-hexynyl phosphoramidite and 3'-alkyne-modifier serinol CPG respectively (Figure 10b).<sup>[9]</sup>

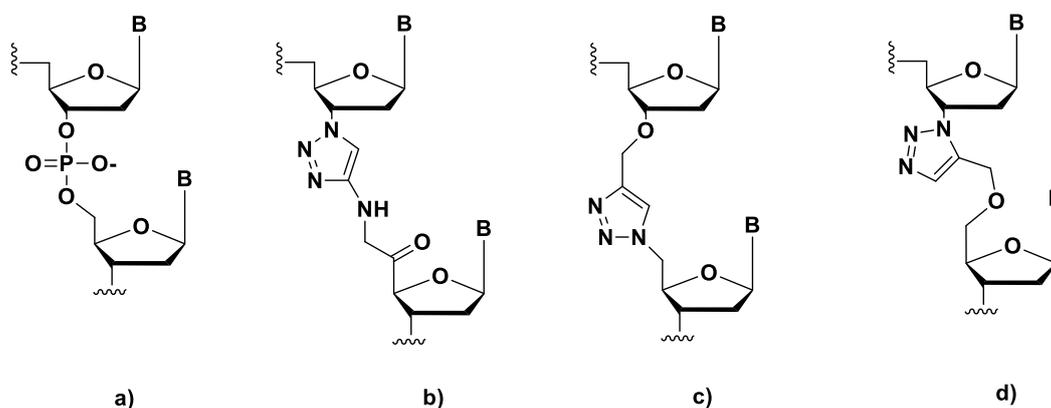
For enzymatic incorporation, 5-Ethynyl-2'-deoxyuridine (EdU) triphosphates are well accepted as substrate by DNA polymerases, in contrast to C8-alkyne-dU, since the longer linkers are attributed to cause oligonucleotide aggregation in H<sub>2</sub>O, thus leading to decreased yields in PCR amplification. Therefore, especially EdU and its derivatives have been applied for metabolic labelling to monitor DNA and RNA synthesis *in vitro* and *in vivo* (animal tissue model).<sup>[9, 38, 47]</sup> In order to circumvent cytotoxic Cu(I) catalyzed click reactions and allow also SPAAC on nucleotides, azide-functionalized uridine derivatives, such as AmdU (5-(azidomethyl)2'-deoxyuridine) for subsequent labelling *via* SPAAC, were investigated (Figure 11a). However, with this approach labelling preferentially occurred on single stranded DNA, very likely due to the bulky cyclooctyne moiety.<sup>[9, 48]</sup>



**Figure 11.** a) Structures of EdU, and AmdU. b) Structure of 5'-azidoguanosine for 5' end labelling of RNA *via* T7 RNA polymerase and 3'-azido-2',3'-dideoxy-ATP/GTP for 3'end labelling of RNA/DNA with PAP and TdT respectively.<sup>[49-50]</sup>

Click functionalized nucleosides and nucleotides have further been used to modify RNA for subsequent CuAAC and SPAAC. Das and co-workers utilized 5'-azidoguanosine to label the 5'-end and 3'-azido-2',3'-dideoxyadenosine to label the 3'-end with T7 RNA polymerase and poly(A)polymerase respectively.<sup>[9, 49]</sup> A similar approach to label the 3'-end of DNA was already introduced in 1986 by Kukhanova *et al.*, using 3'-azido-2',3'-dideoxyguanosine in combination with TdT (terminal deoxynucleotidyl transferase) (Figure 11b).<sup>[50-51]</sup> Upon labelling the 3'-end

of DNA or RNA with a clickable functionality, it is possible to utilize subsequent triazole formation by CuAAC, for chemically ligated nucleic acids. In 2009, the Brown laboratory started to investigate the biocompatibility of triazole linked DNA (3'-azido-dT clicked to 5'-propargylamido-dT), towards its acceptance of polymerases (Figure 12b). Since one T base was skipped during PCR, a different triazole linkage (5'-azide clicked to 3'-propargyl modified oligonucleotides) was introduced, achieving successful amplification of up to two triazole backbone modifications per template by DNA polymerases that do not possess proof reading activity (Figure 12c).<sup>[52-53]</sup> Further, DNA bearing these triazole backbones was successfully transcribed into RNA by T7 RNA polymerase.<sup>[54]</sup> In 2021, the Brown laboratory reported a new 1,5-disubstituted triazole (prepared *via* RuAAC), better mimicking the phosphodiester backbone for enhanced polymerase compatibility. (Figure 12d).



**Figure 12.** a) Phosphodiester linkage (5 bonds) b) First generation triazole linkage (7 bonds) c) 2<sup>nd</sup> Generation triazole linkage (7 bonds), successfully accepted by DNA polymerase and RNA polymerase. d) New 1,5-triazole linkage introduced by Brown *et al.*, with enhanced polymerase compatibility (6 bonds).<sup>[24]</sup>

## 1.3 Application of Click-Functionalized Nucleic Acids

### 1.3.1 Therapeutic Applications

Click functionalized nucleic acid building blocks stimulated interest for various applications, including therapeutic purposes. Within this context, messenger RNA (mRNA) based vaccines display a notable example which experienced a breakthrough during the SARS-CoV-2 pandemic. Katalin Karikó and Drew Weissman were awarded with the Nobel Prize 2023 for Physiology and Medicine, for their discoveries concerning nucleotide base modifications, leading to the development of mRNA based vaccines against SARS-CoV-2.<sup>[55]</sup> Their key invention is based on the replacement of uridine with pseudouridine ( $\Psi$ ) leading to decreased immunogenicity while increasing translation capacity and biological stability, which was further improved by Andries *et al.*, *via* the incorporation of *N*<sup>1</sup>-methyl-pseudouridine ( $m^1\Psi$ ), providing superior protein expression and reduced immunogenicity.<sup>[56-57]</sup> In 2020, Croce and co-workers introduced the incorporation of 5-ethynyl-uridine-trisphosphate (EUTP) using T7 RNA polymerase during *in vitro* transcription (IVT), followed by CuAAC with a fluorescent dye in order to track the alkyne-modified mRNA inside cells, whereas protein expression (eGFP) was observed comparably to unmodified mRNA. This approach was further combined with 3'-terminal azide labelling *via* yeast poly(A) polymerase (yPAP) and 3'-azido-2',3'-dideoxyadenosine (AzddATP), allowing the chemical attachment of any molecule of interest, such as targeting agent.<sup>[58]</sup> Encouraged by these results, the baseclick GmbH in 2020 began the development of an mRNA based vaccine against SARS-CoV-2, applying click chemistry for the attachment of a targeting agent at the 3'-end of the N-protein encoding mRNA (section 4.4.1). Maassen and co-workers further investigated the replacement of uridine in mRNA by EU for its potential usage in RNA-based vaccines, revealing that EU-modified mRNA exhibits lower expression levels compared to  $\Psi$  and  $m^1\Psi$ -modified mRNA. However, its stability and low immunogenicity are comparable to other uridine analogues, emphasizing the potential of EU for RNA based vaccines.<sup>[59]</sup>

Another application of therapeutic nucleic acids in which click chemistry is involved are RNAi (RNA interference) based drugs, such as small interfering RNAs (siRNAs). Similar to the click modified mRNA, click chemistry is applied during siRNA development for targeted delivery.<sup>[60]</sup> While one approach is to mannosylate nanoparticles for targeted delivery of the encapsulated

siRNA *via* click chemistry, therapeutic siRNA can be clicked to a receptor ligand using Cu(I)-catalyzed click-chemistry, as reported by the Carell laboratories.<sup>[61-63]</sup> Their approach of ACE2-binding peptide-conjugated siRNA against SARS-CoV-2, whereas the N<sub>3</sub>-modified peptide is attached to the alkyne-bearing dT unit at the 2<sup>nd</sup> position at the 3'-end of the siRNA, showed promising results and represents exceptional opportunities for the advancement of antiviral treatments.<sup>[64]</sup> Moreover, click chemistry allows to modify siRNA with other moieties, such as antibodies in a straightforward manner.<sup>[65]</sup>

Similarly, antisense oligonucleotides (ASOs) can be attached to antibodies *via* click chemistry, whereas the ASO is delivered as a duplex and the carrier strand is bearing the click-ligated antibody, which, upon cellular entry is degraded and releases the ASO.<sup>[65-66]</sup> Further, triazole backbones have been investigated towards their use in ASOs, whereas a 1,4 disubstituted-triazole, generated by CuAAC, exhibited the highest proficiency in adopting the A-conformation, necessary for RNA recognition, and thus establishing it as a promising candidate for integration into antisense oligonucleotides.<sup>[9, 67]</sup>

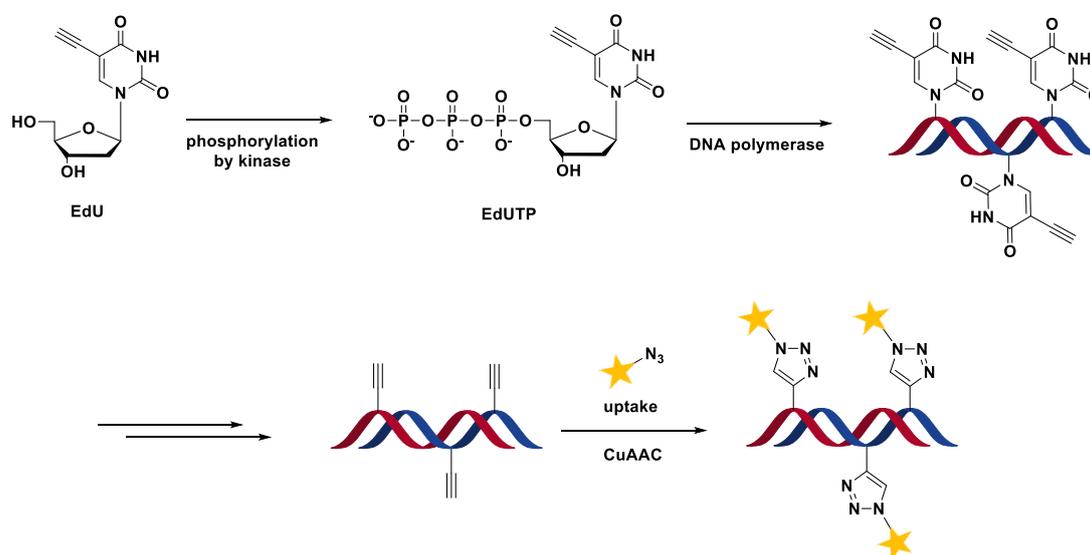
In another recent study, Beck and co-workers introduced a trimannose-conjugated microRNA oligonucleotide drug, directed against miR21 (LNA-antimiR-21), which is supposed to interfere with pathological macrophages in acute inflammatory lung disease. Their findings suggest that coupling with trimannose using click chemistry, enhances the efficient and selective delivery of inhaled oligonucleotides to pulmonary macrophages.<sup>[68]</sup>

With the advent of novel clickable nucleosides, nucleotides, and modification reagents for oligonucleotides, focusing on regulatory RNAs emerges as a promising direction, potentially leading to significant advancements in the realm of synthetic gene therapeutics.<sup>[60]</sup>

### 1.3.2 Diagnostic Applications

#### 1.3.2.1 Metabolic Labelling

In terms of diagnostic purposes, labelling of click-modified nucleic acids with fluorescent dyes still represents the main application. One example, that was previously mentioned and laid the foundation for the baseclick GmbH is the EdU-based cell proliferation analysis. Herein, EdU gets phosphorylated inside cells and further incorporated into cellular DNA by polymerases during replication. The exposed alkynes can further undergo labelling *via* CuAAC with azide-bearing fluorescent dyes on fixed cells for subsequent imaging (Figure 13).<sup>[9, 38]</sup>



**Figure 13.** Schematic representation of the EdU based cell proliferation technology. First EdU gets phosphorylated into cells, and subsequently EdUTP is integrated into the genome by DNA polymerases. The alkynes can then be labelled *via* CuAAC with azide-bearing dyes.<sup>[9, 38]</sup> ★ = fluorescent dye.

### 1.3.2.2 Next and Third Generation Sequencing

Moreover, next-generation and third-generation sequencing (NGS, TGS) represent emerging applications for click-modified nucleic acids. Especially, the emerging field of third-generation sequencing platforms, capable of producing long reads (multi-kilobases in length) has been using click chemistry for various steps during library preparation and sequencing.<sup>[69]</sup>

Regarding library preparation, Routh and colleagues introduced a PAC-seq (poly(A) Click-Seq) method, which employs reverse transcription-PCR (RT-PCR) in the presence of 3'-azido-ddNTPs, such that stochastic chain termination takes place upstream of the poly(dT)-primed 3'-UTR, enabling 3'-end sequencing without RNA enrichment or fragmentation.<sup>[70]</sup>

In terms of the sequencing itself, one of the leading TGS platforms, Pacific Biosciences (PacBio) utilizes click chemistry for the preparation of their fluorophore labelled nucleotides. Herein, they not only use CuAAC but also SPAAC for the preparation of dye-modified nucleotide reagents.<sup>[71-72]</sup> Within this context, the synthesis of  $\gamma$ -labeled nucleoside 5'-triphosphates, bearing a clickable functionality, reported by Serdjukov *et al.*, as well as the work from the Marx laboratory, were definitely pivotal for advancements regarding phosphate modified triphosphates finding nowadays usage in TGS.<sup>[73]</sup>

The emerging number of publications and patents, covering the synthesis and usage of click modified nucleotides for sequencing applications, as well as advancements in protein and epigenetic sequencing involving click chemistry, underlines the value of this biorthogonal reaction for diagnostic applications.<sup>[74-77]</sup>

## 2. Aim of this Thesis

The objective of this doctoral thesis is to conduct thorough examinations of click chemistry on nucleic acids, with a specific focus on its application in diagnostics and therapeutics.

The primary goal of this thesis is directed to the field of Next and Third generation sequencing, namely the development of a click-chemistry based library preparation method. By incorporating straightforward Cu(I) catalyzed azide-alkyne cycloaddition click chemistry on complementary DNA (cDNA) during the library preparation workflow, we aim to generate superior and less error prone results for full-length mRNA sequencing. To assess the applicability of the anticipated technology, we aim to design a method, that can be performed in straightforward manner in all kind of laboratories, without sophisticated setup and the need of special equipment and chemicals, such as inert gas. Furthermore, the goal is to make this technology accessible for subsequent sequencing on all dominantly used sequencing methods, such as Oxford Nanopore Technologies (ONT) or Pacific Bioscience (PacBio).

The secondary goal of the work presented here is to investigate the utility of  $\gamma$ -phosphate modified ATP analogues, bearing a clickable functionality as substrates for a polynucleotide kinase, namely T4 PNK. Specifically, the aim is to develop a novel method to label the 5'-end of nucleic acids with a clickable functionality in order to achieve subsequent straightforward attachment of various molecules of interest, such as targeting agents or sequencing adaptors, by using click chemistry. This approach would afford the synthesis of different, click-functionalized ATP analogues and the exploration of their utilization towards T4 PNK on a nucleic acid model system.

The third objective is to develop a click modified Cu(II)-dependent nucleic acid based artificial ribonuclease, targeting a conserved region of the SARS-CoV-2 genomic RNA in order prevent the viral replications.

The fourth goal is to improve mRNA vaccine developments, applying copper free click chemistry for quantitatively modifying therapeutic mRNA with targeting moieties or using click chemistry for the preparation of circular RNA.

### 3. Published Works

#### 3.1. First Publication

**TITLE:** Click Chemistry Enables Rapid Amplification of Full-Length Reverse Transcripts for Long-Read Third Generation Sequencing

**AUTHORS:** Eva S. Schöneegger<sup>+</sup>, Antony Crisp<sup>+</sup>, Markus Müller, Jessica Fertl, Sascha Serdjukow, Stefano Croce, Michael Kollaschinski, Thomas Carell, and Thomas Frischmuth\*

*+ authors contributed equally (shared 1<sup>st</sup> authorship). \* corresponding author.*

**JOURNAL, YEAR, VOLUME AND PAGE NO.:**

Bioconjugate Chemistry **2022**, 33 (10), 1789-1795.

#### **SUMMARY:**

The most commonly applied technology for preparation of full-length cDNA libraries from total RNA involves a template-switching (TS) reaction. To overcome concerns and limitations of this widely used strategy, a novel click chemistry based method for the generation and amplification of full-length cDNA libraries from total RNA is introduced. This click chemistry-based library preparation protocol avoids not only random priming, but also stochastic cDNA termination and therefore allows the amplification of full-length reverse transcripts, which could not be enabled by previous click-chemistry related RNA sequencing methods. The introduced PCR primers have superior impact on the read-through compatibility of the 1,4-disubstituted 1,2,3-triazole-containing cDNA. This is enabled by two overhanging 3'-nucleotides and increases the possible insert size in comparison to the PAC -seq. technique (poly(A)-ClickSeq). Moreover, the novel click chemistry based library preparation method gave comparable results to a commercially available cDNA-PCR sequencing kit, developed by Oxford Nanopore Sequencing. Thus, this novel technology could contribute to advancements in fusion transcript and mRNA splicing variant investigations.

**PERSONAL CONTRIBUTIONS:**

- Synthesis of mRNA and cDNA as depicted in the publication.
- Performed 3'-end labelling of cDNA (azide elongation), CuAAC reactions for ligation of the 3'-azido-labeled cDNA with the alkyne adapter and PCR amplifications.
- Conducted agarose gel electrophoresis analysis.
- Performed all sequencing experiments using Oxford Nanopore MinION device.
- Composed, edited and revised the manuscript and supporting information together with all authors.

**PERMISION AND LICENSING**

Reprinted with permission from E. S. Schönegger, A. Crisp, M. Müller, J. Fertl, S. Serdjukow, S. Croce, M. Kollaschinski, T. Carell, T. Frischmuth, *Bioconjugate Chemistry* **2022**, *33*, 1789-1795.  
<https://doi.org/10.1021/acs.bioconjchem.2c00353> © 2022 American Chemical Society

# Click Chemistry Enables Rapid Amplification of Full-Length Reverse Transcripts for Long-Read Third Generation Sequencing

Eva S. Schönegger,<sup>#</sup> Antony Crisp,<sup>#</sup> Markus Müller, Jessica Fertl, Sascha Serdjukow, Stefano Croce, Michael Kollaschinski, Thomas Carell, and Thomas Frischmuth\*



Cite This: *Bioconjugate Chem.* 2022, 33, 1789–1795



Read Online

ACCESS |



Metrics & More

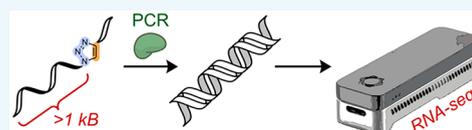


Article Recommendations



Supporting Information

**ABSTRACT:** Here we describe the development of a novel click chemistry-based method for the generation and amplification of full-length cDNA libraries from total RNA, while avoiding the need for problematic template-switching (TS) reactions. Compared with prior efforts, our method involves neither random priming nor stochastic cDNA termination, thus enabling amplification of transcripts that were previously inaccessible via related click chemistry-based RNA sequencing techniques. A key modification involving the use of PCR primers containing two overhanging 3'-nucleotides substantially improved the read-through compatibility of the 1,4-disubstituted 1,2,3-triazole-containing cDNA, where such modifications typically hinder amplification. This allowed us to more than double the possible insert size compared with the state-of-the-art click chemistry-based technique, PAC-seq. Furthermore, our method performed on par with a commercially available PCR-cDNA RNA sequencing kit, as determined by Oxford Nanopore sequencing. Given the known advantages of PAC-seq, namely, suppression of PCR artifacts, we anticipate that our contribution could enable diverse applications including improved analyses of mRNA splicing variants and fusion transcripts.



Beyond its initial synthesis, eukaryotic pre-mRNA undergoes a multitude of co- and post-transcriptional processing events that dictate how and when it will be translated by the cellular apparatus. Our understanding of these processes has matured largely thanks to advancements in RNA sequencing (RNA-seq) technologies.<sup>1</sup> In order to ascertain the function of a particular gene transcript, capture and sequencing of all possible variants are required, since the biological function of any particular spliced and polyadenylated form may be distinct.<sup>2</sup> Due to technical limitations associated with library generation, however, post-transcriptional regulatory events can often go undetected.<sup>3</sup> Given that the majority of mammalian transcripts are between 1 and 2 kb in length,<sup>4,5</sup> long-read technologies are crucial when it comes to disentangling the entire spectrum of mRNA isoforms.<sup>5</sup> The most widely exploited strategy for cDNA library preparation from total RNA involves the use of methods derived from the Smart-seq2 protocol.<sup>6,7</sup> As with many other approaches, Smart-seq2 involves a template switching (TS) reaction whereby primer-specific 3'-adapter sequences can be appended to nascent cDNA ends.<sup>8</sup> This process begins with an initial 3'-extension with three protruding 2'-deoxycytidine nucleotides, which may then subsequently hybridize to a reverse-complementary rGrGrG-containing template-switching oligo (TSO). This allows the reverse transcriptase to shift between the DNA and RNA templates and to copy the entire TSO sequence to the 3'-end of the resulting cDNA.<sup>5</sup> While groundbreaking in its original implementation,<sup>9</sup> template switching introduces a host of concerns, including inefficient ligation,<sup>10,11</sup> over-representation of 5'-guanosine-containing

sequences,<sup>12</sup> as well as major artifacts and false alternative transcripts.<sup>13</sup>

Inspired by earlier work from Routh and co-workers,<sup>14,15</sup> we became interested in developing a cDNA library preparation method that not only harnessed the established benefits of click chemistry, but was also compatible with long-read third generation sequencing. Compared with enzymatic processes, Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC) reactions have several distinct advantages. Namely, CuAAC reactions are straightforward to execute, are stereospecific, are often quantitative in yield, and result in few to no byproducts.<sup>16</sup> With these advantages in mind, Routh and co-workers reported a click chemistry-based approach for the capture and elucidation of 3'UTR/poly(A) tail junctions via short-read sequencing.<sup>14,15</sup> Their protocol titled PAC-seq involved RT-PCR in the presence of 3'-azido-ddNTPs, such that stochastic chain termination always occurs upstream of the poly(dT)-primed 3'UTR. While ideally suited for sensitive and specific amplification of the first 200–400 nucleotides upstream from polyadenylated sites, PAC-seq does not yield coverage of many full-length transcripts, likely owing to the limited polymerase read-through efficiency of various triazole-

Received: July 28, 2022

Revised: September 21, 2022

Published: September 26, 2022

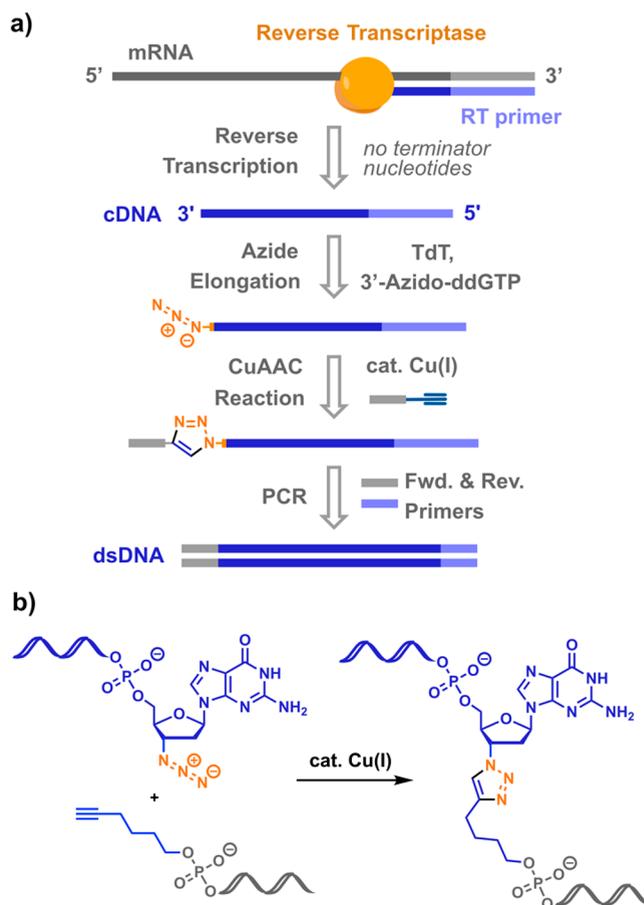


containing ssDNA templates.<sup>17</sup> We thus sought to establish an enhanced method that would not only target the 3'-ends of polyadenylated RNAs, but also the entirety of any given transcript. We envisaged that this could be achieved via the introduction of a subsequent and separate terminal deoxynucleotidyl transferase (TdT)-catalyzed ligation reaction after reverse transcription to incorporate the relevant 3'-azido-ddNTP, as has previously been exploited within our laboratories.<sup>18,19</sup> The advantage in doing so would allow the entire transcript to be sequenced as a single read, harnessing the full capacity of Third Generation Sequencing technology, whereas stochastic termination cDNA synthesis can only provide reads of a certain predetermined, average length. Compared with other long-read library preparation methods including but not limited to Smart-Seq2<sup>10</sup> or those described in commercially available kits from Oxford Nanopore Technologies, the potentially problematic one-step RT-PCR is avoided. Our method is also thus expected to exhibit the other ostensible benefits of ClickSeq methods such as circumvention of artifactual chimera formation and of certain TS-derived sequence biases.<sup>14,20</sup>

## RESULTS AND DISCUSSION

**Development of cDNA Library Preparation Using CuAAC.** With the previous literature precedent<sup>14,15,21,22</sup> as well as our planned improvements in mind, we designed our cDNA library preparation method as shown in Figure 1 a. Reverse transcription, using eGFP mRNA as template (Figure S6), is initially performed under standard conditions, employing an anchored poly(dT)-containing primer to ensure full coverage of the 3' UTR. After cDNA synthesis and digestion of the RNA template, nascent cDNAs are treated with a buffered solution containing 3'-azido-2',3'-dideoxyguanosine-5'-triphosphate in the presence of TdT, which results in addition of a single azido-modified nucleotide at the 3'-cDNA end. The mixture is then purified via a silica-membrane-based purification system to provide the single-stranded 3'-azido-terminated cDNAs as a concentrated solution in unbuffered-H<sub>2</sub>O. The click ligation is accomplished using Reactor 25, a solid, heterogeneous Cu(I) catalyst system developed at baseclick GmbH in the presence of THPTA, MgCl<sub>2</sub>, and an adapter oligodeoxynucleotide (ODN) containing a 5'-(5-hexynyl) phosphate moiety (Table S1), the sequence of which was closely derived from that which was designed by Routh and co-workers.<sup>14</sup> This reaction provides the 3'-adapter-ligated cDNA that will serve as the template for further PCR amplification (Figure 1 b).

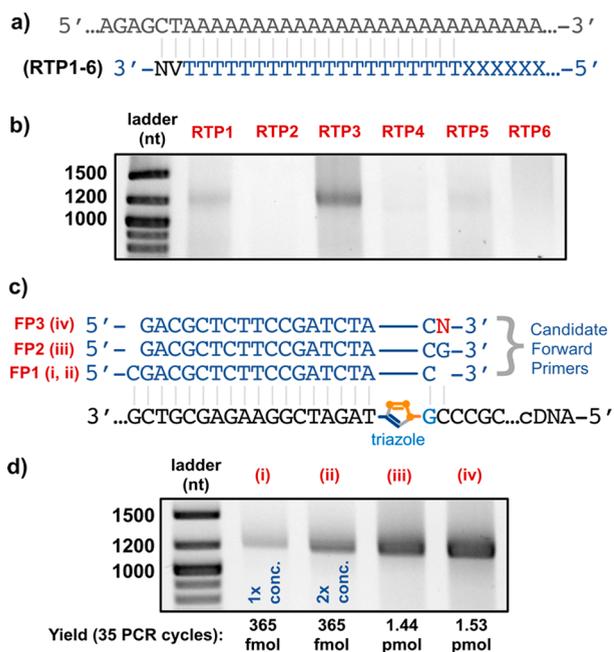
Given that biocompatible triazole linkers are known to have a diminishing effect on the yields of enzymatic reactions that copy through the unnatural backbone,<sup>23</sup> certain considerations to primer design were made to ensure the highest possible yield during the PCR amplification. First, a PCR forward primer was designed such that a single, overhanging deoxycytidine nucleotide will hybridize to the 3'-azido-2',3'-dideoxyguanosine nucleotide of the reverse-complementary cDNA strand (Figure 2 a). This design ensures bridging of the triazole backbone with the primer and thus avoiding primer extension (or replication) across the potentially unfavorable linkage. The design also enables primer extension of the sense strand irrespective of the sequence context. To properly assess the capacity of this primer to facilitate PCR, short amplifications were set up involving a sequence specific reverse primer (Figure 2 b; Table S1). RTP3 was identified as the best-



**Figure 1.** (a) Schematic overview describing how dsDNA libraries are prepared from extracted total-RNA or isolated mRNA sequences. (b) The copper-catalyzed azide–alkyne cycloaddition (CuAAC) reaction of 3'-azido labeled cDNAs to provide the sequence-specific template required for PCR amplification of the ssDNA library.

performing RT primer and thus 3'-clicked cDNA, obtained using RTP3, was amplified using each of three candidate forward primers (FP1–3) in combination with reverse primer RP3 (Figure 2 c) to further improve the yield and reproducibility of the PCR. Agarose-gel visualization and quantification using a Qubit Fluorometer indicated that yields are substantially enhanced by the incorporation of an extra, fully degenerate (N) nucleotide at the 3'-end of the forward primer (Figure 2 d). Primer and adapter sequences are indicated in Table S1.

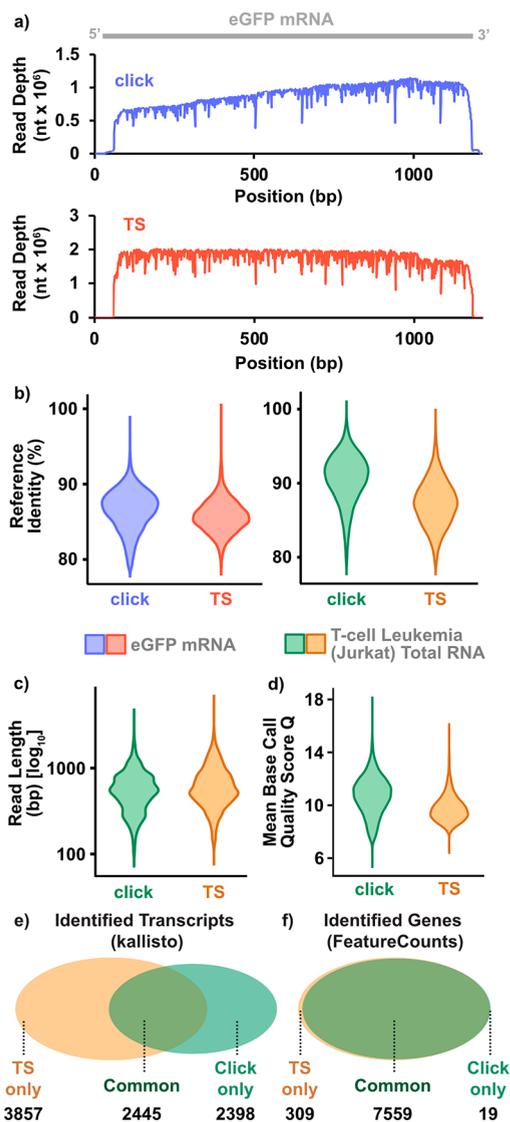
**Evaluation of Click Chemistry-Based dsDNA Libraries via ONT Third Generation Sequencing.** Having now demonstrated that triazole-conjugated ssDNA products of greater than 1 kB can successfully serve as templates for PCR, we next went about generating dsDNA libraries for sequencing with the Oxford Nanopore Technologies (ONT) platform. To this end, eGFP-encoding dsDNA products derived from clicked-cDNA were enzymatically repaired, and subsequently ligated to sequencing adaptors using a commercially available ligation-based sequencing kit designed for multiplexing samples (ONT Ligation sequencing gDNA kit: SQK-LSK109). The library was then sequenced using nanopore MinION device for 24 h. An analogous experiment was subsequently conducted, in which the initial dsDNA library was instead prepared using a commercially available kit that ostensibly uses template switching (TS) with supplied adapter



**Figure 2.** Optimization of RT- and PCR-primers allows the amplification of an *in vitro* transcribed model 1.22 kB mRNA coding for enhanced green fluorescent protein (eGFP), which was first reverse transcribed using six candidate RT-primers (a) (RTP1 – RTP6) and then labeled using 3'-azido-2',3'-ddGTP in combination with the enzyme TdT. After the Cu(I)-catalyzed click-ligation of the resulting 3'-azido-cDNAs with the alkyne-adapter AAI, the material was amplified by PCR. (b) Crude PCR products originating from each RT-primer candidate were then visualized by agarose gel-electrophoresis. (c) 3'-Clicked cDNA obtained using the best-performing RT primer (RTP3) was amplified using each of three candidate forward primers (FP1–3) in combination with reverse primer RP3. (d) Agarose-gel visualization and quantification using a Qubit Fluorometer indicated that yields are substantially enhanced by the incorporation of an extra, fully degenerate (N) nucleotide at the 3'-end of the forward primer. Primer and adapter sequences are indicated in Table S1.

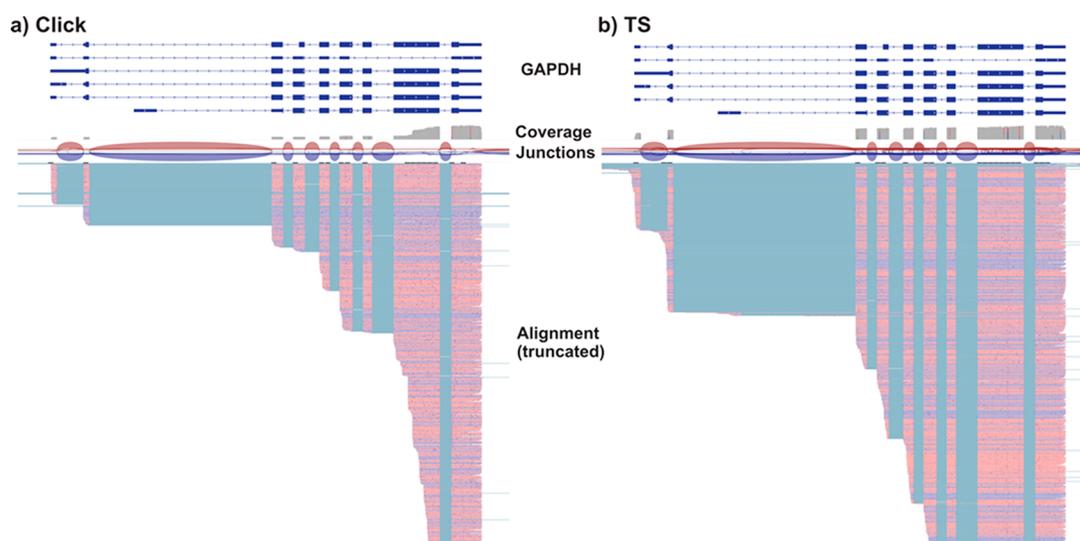
ODNs. The sequencing run was also performed for 24 h. Coverage depth diagrams associated with the two experiments are shown in Figure 3a. Here, the relative read depths indicated that our click chemistry-based library preparation leads to comparable results to those obtained using the standard library preparation, except with a slight decline in coverage toward the 5'-end. Since the CuAAC reaction occurs irrespective of sequence context, this might suggest that the click protocol leads to successful adapter incorporation even in truncated cDNAs that would not ordinarily be ligated in the TS protocol. We attribute the slightly lower total read number in the click-library to normal variations in flow cell performance. Moreover, while 1363 pores yielded sequence data in the TS run, only 942 pores were productive for the Click library (active pore count was measured during the flow cell check via MinKNOW software).

To demonstrate the applicability of our method toward whole transcriptome RNA-seq, dsDNA libraries derived from clicked-cDNA were prepared using T cell leukemia (Jurkat) total RNA. Following enzymatic end-repair and sequencing adapter ligation, as with our eGFP model system, nanopore sequencing was performed for 48 h run time. Again, an analogous experiment was conducted, in which the sequencing



**Figure 3.** Performance of the Click-libraries in nanopore sequencing. (a) Coverage depth (read numbers) on the eGFP model transcript of two independent runs on a MinION device (blue: click library, crimson: TS library). (b) Comparison of mapping quality (percent reference identity) of the eGFP model libraries and Jurkat total RNA libraries, prepared by click chemistry (click) or standard protocol (TS). (c) Read-length distribution of the two Jurkat total RNA libraries. (d) PHRED base call quality score comparison of click vs standard library. (e) Transcripts identified with the Kallisto algorithm.<sup>24</sup> Cut-off at >10 transcripts per million (tpm). (f) Genes identified from read-counts that were mapped to exons of the human genome (minimap2). Reads counted and normalized with *featureCounts*.<sup>25</sup> Cut-off at >10 counts per million (cpm).

library was generated by a commercially available kit using template switching (TS), and the library obtained was sequenced under the same conditions as the CuAAC-based sample for a total of 48 h. The mapping quality of the individual sequencing experiments is shown in Figure 3b. Notably, we observed a higher percent reference identity for the click-generated libraries in comparison to standard library preparation. We speculate that this effect may at least partially be attributed to 5'-guanosine-containing sequences present only in the TS protocol (Figure S9). Read lengths (Figure 3c) of the CuAAC based library preparation were comparable to



**Figure 4.** Mapping example, showing the GAPDH gene on chromosome 12 for two Jurkat total RNA libraries sequenced on a MinION nanopore sequencer. Mapping was performed with minimap2 against the GRCh38 reference genome. (a) The library that was prepared using the click chemistry protocol. (b) The library that was prepared with the standard template switching procedure (TS). The images were prepared with IGV 2.12.2. Reads in the forward direction (or 5' to 3', with respect to the mRNA) are colored red, and those in the reverse direction are colored blue. Splice junctions connecting exons are represented by arcs.

the standard library preparation, whereas PHRED base call quality scores were slightly higher in the case of the Click-chemistry method (Figure 3d). The slight disparity in quality scores can perhaps be attributed to variation in the fidelity of the polymerases employed in each experiment. In the case of the proprietary TS kit, the identities of the polymerases were undisclosed, precluding further investigation into this effect. Finally, to determine and compare transcript abundance from the nanopore data, we employed the mapping-independent Kallisto algorithm developed by Bray and colleagues (Figure 3e).<sup>24</sup> To our surprise, we find only about 50% overlap with the TS library. When we collapse the identified transcripts to gene level, the overlap increases, indicating a problem in transcript identification.

To gain further insight into this problem, we mapped the reads to the human genome. An illustrative example of full-length transcript mapping is shown in Figure 4. Specifically, ONT reads derived from Jurkat total RNA were mapped to the GAPDH gene on chromosome 12 using data derived from both (a) the click chemistry method and (b) the TS method. As anticipated, the data show that GAPDH exons are well represented using our click-chemistry based library preparation, with a slight bias toward the 3'-end, as compared with the TS-based kit.

The outcome that the mapping showed a larger proportion of shorter reads in the Click library compared to the TS library was not unexpected, given that this was also observed in the eGFP library (Figure 3a). We attribute this effect to our methodology, which enriches more truncated reads compared with TS protocols, since the introduction of azido-nucleotides does not depend on the RT reaction reaching the 5'-end of a given mRNA template. This is further illustrated by the fact that both methods perform better when tallying the sum of identified genes rather than transcripts. Since the kallisto algorithm is still based on transcript identification, we next applied the *featureCounts* algorithm from the Rsubread package<sup>25</sup> to the mapped reads. Indeed, when applying this

algorithm, we observed a very good overlap between the two libraries (Figure 3f).

## CONCLUSION

We here disclose a novel click chemistry-based method for the preparation of long-read sequencing libraries from polyadenylated RNA. Click chemistry-based dsDNA libraries for Third Generation Sequencing were prepared by 3' azido-labeling of reverse transcribed cDNA via an anchored poly(dT) primer, followed by a CuAAC and subsequent PCR amplification. All individual steps of the introduced library preparation are reproducible, high yielding, and easy to perform. By the incorporation of an extra, quadruple degenerated nucleotide at the 3'-end of the forward primer, the quality of the PCR was enhanced significantly, allowing polymerase-mediated amplification of lengthy triazole-containing ssDNA templates.

We directly compared Nanopore sequencing experiments of the two libraries (eGFP and Jurkat), generated via click chemistry-based and standard protocol (based on template switching reaction) library preparation, respectively. The generated sequencing data demonstrate that the performance of the click-chemistry based library in both amplicon and transcriptome sequencing can compete with methods that involve template switching. Unbiased addition of sequencing adapters may further facilitate RNA-seq protocols that rely on blockage of reverse transcription like in SHAPE-seq or sequencing of RNA modifications (e.g., m<sup>1</sup>A or pseudouridine), which prohibit the application of template switching.<sup>26–29</sup>

In conclusion, our novel library preparation method enables the reliable generation and amplification of full-length reverse transcripts for long read sequencing while avoiding templated ligation, artifactual chimeras, and over-representation of 5'-guanosine-containing sequences.<sup>14</sup> Compared with the established protocol PAC-seq,<sup>21,22</sup> mRNA transcripts longer than 1 kb are now within reach. Our method also offers an attractive alternative to Smart-Seq2,<sup>8</sup> potentially forming the basis for future development of more sophisticated workflows such as

single-cell and long-read amplicon sequencing.<sup>30</sup> Given the modular-nature of click chemistry, an extremely diverse range of applications is in principle possible. Our library preparation could further be applied to the analysis of differential gene expression and quantification of splicing variants as well as fusion transcripts.

## EXPERIMENTAL SECTION

**Custom Oligonucleotides.** Short oligonucleotides and primers used in this study were provided by baseclick GmbH (Neuried, Germany) and were synthesized using standard phosphoramidite chemistry. HPLC-purified products were stored at  $-20\text{ }^{\circ}\text{C}$  as lyophilized solids.

**mRNA Synthesis.** In order to generate cDNA libraries for Third Generation Sequencing, mRNAs encoding for the enhanced green fluorescent protein (eGFP) and T cell leukemia (Jurkat) total RNA were first prepared or purchased as model systems. Jurkat Total RNA was purchased from ThermoFisher Scientific, while the desired mRNA constructs of eGFP (Figure S5) were prepared by *in vitro* transcription from a linearized DNA template containing a T7 promoter and a 120mer poly(A) tail according to a modified version of a previously established protocol from Croce and co-workers (Figure S4).<sup>18</sup> Reaction mixtures were prepared with the TranscriptAid T7 High Yield Transcription Kit (ThermoFisher Scientific) by mixing the following components in a 200  $\mu\text{L}$  PCR tube: linear plasmid DNA (1  $\mu\text{g}$ ); 8  $\mu\text{L}$  NTP mix (2.5 $\times$ ; 100 mM ATP/CTP/GTP/UTP solutions); 4  $\mu\text{L}$  TranscriptAid Reaction Buffer (5 $\times$ ); 2  $\mu\text{L}$  TranscriptAid Enzyme Mix and nuclease-free  $\text{H}_2\text{O}$  to achieve a final volume of 20  $\mu\text{L}$ . The samples were incubated for 2 h at  $37\text{ }^{\circ}\text{C}$ , cooled to  $4\text{ }^{\circ}\text{C}$  for 3 min and purified with the QIAquick PCR Purification Kit from Qiagen. The concentrations were measured with the Qubit Fluorometer and the samples quality was analyzed via agarose gel electrophoresis.

**First Strand cDNA Synthesis.** Prior to reverse transcription (RT), the poly(dT) primers (RTP1-RTP6) were hybridized to the template by mixing the following components in a 200  $\mu\text{L}$  PCR microtube: 1  $\mu\text{L}$  of eGFP mRNA (600–650 ng) or Jurkat total RNA (1–2  $\mu\text{g}$ ); 1  $\mu\text{L}$  dNTP Mix (10 mM; New England Biolabs); 1  $\mu\text{L}$  (RTP1-RTP6; 100  $\mu\text{M}$ ), and 10  $\mu\text{L}$  nuclease-free  $\text{H}_2\text{O}$  to achieve a final volume of 13  $\mu\text{L}$ . After mixing with a pipet, the mixture was briefly centrifuged and transferred to a thermal cycling block with a lid temperature of  $105\text{ }^{\circ}\text{C}$ . The samples were then heated to  $65\text{ }^{\circ}\text{C}$  for 5 min and cooled to  $4\text{ }^{\circ}\text{C}$  for 3 min. Afterward, 4  $\mu\text{L}$  of SuperScript IV 5 $\times$  Buffer, 1  $\mu\text{L}$  of DTT (100 mM), 1  $\mu\text{L}$  of SuperScript IV Reverse Transcriptase (200 U/ $\mu\text{L}$ ), and 1  $\mu\text{L}$  of RNaseOUT Recombinant Ribonuclease Inhibitor (40 U/ $\mu\text{L}$ ) were added to reach a final volume of 20  $\mu\text{L}$ . The resulting samples were mixed and briefly centrifuged, and the following temperature program for the RT was performed: 20 min at  $50\text{ }^{\circ}\text{C}$ , 10 min at  $80\text{ }^{\circ}\text{C}$ , cool to  $4\text{ }^{\circ}\text{C}$ , and hold for 3 min. SuperScript IV 5 $\times$  Buffer, DTT, and SuperScript IV Reverse Transcriptase and RNaseOUT Recombinant Ribonuclease Inhibitor were purchased from ThermoFisher Scientific.

For the following mRNA digestion, to each reaction mixture from the previous step were added to the following components: 3  $\mu\text{L}$  RNase H Reaction Buffer (10 $\times$ ), 1.4  $\mu\text{L}$  RNase H (5 U/ $\mu\text{L}$ ), 1  $\mu\text{L}$  RNase A (10 mg/ $\mu\text{L}$ ), and 4.6  $\mu\text{L}$  Nuclease-free  $\text{H}_2\text{O}$  to achieve a final volume of 30  $\mu\text{L}$ . The digestion was performed as follows: 30 min at  $37\text{ }^{\circ}\text{C}$ , 15 min at

$65\text{ }^{\circ}\text{C}$ , cool at  $4\text{ }^{\circ}\text{C}$ , and hold for 3 min. The digested cDNAs were then purified using a QIAquick PCR Purification Kit (Qiagen). Elution was performed using nuclease-free  $\text{H}_2\text{O}$ . RNase H Reaction Buffer (10 $\times$ ) and RNase H were purchased from New England Biolabs. RNase A was purchased from ThermoFisher Scientific.

**Azide Elongation and CuAAC.** The samples for the azide elongation were prepared by mixing the following components in a 200  $\mu\text{L}$  PCR tube: 15 to 17  $\mu\text{L}$  purified cDNA (eGFP: 1000–1300 ng; Jurkat: 900–950 ng); 1  $\mu\text{L}$  3'-Azido-2',3'-ddGTP (10 mM; BASECLICK GmbH); 2  $\mu\text{L}$  TdT (20 U/ $\mu\text{L}$ ); 5  $\mu\text{L}$  TdT Reaction Buffer (5 $\times$ ), and if necessary, sufficient nuclease-free  $\text{H}_2\text{O}$  to achieve a final volume of 25  $\mu\text{L}$ . Azide elongation was performed by incubating the sample for 1.5 h at  $37\text{ }^{\circ}\text{C}$ , then cooling to  $4\text{ }^{\circ}\text{C}$ , and holding for 3 min. The samples were purified as described above, again using nuclease-free  $\text{H}_2\text{O}$  for elution. TdT and TdT Reaction Buffer (5 $\times$ ) was purchased from ThermoFisher Scientific.

CuAAC reactions to ligate the 3'-azido-labeled cDNA with the alkyne adapter (AA1, see Table S1) were prepared by mixing the following components in 1.5 mL tube: a solution containing the single-stranded  $\text{N}_3$ -labeled cDNA (eGFP: 440–460 ng; Jurkat: 450–650 ng) obtained from the previous step in nuclease free  $\text{H}_2\text{O}$  (9 to 28  $\mu\text{L}$ ) was combined by mixing with an appropriate volume of baseclick's proprietary 5 $\times$  activator solution to obtain a 1 $\times$  final concentrated solution. Afterward 2 pellets of baseclick's proprietary heterogeneous reactor catalyst system, 0.5  $\mu\text{L}$  AA1 (100  $\mu\text{M}$ ), and if necessary, nuclease-free  $\text{H}_2\text{O}$ , were added and mixed (final volume 12.5–40  $\mu\text{L}$ ). The click reactions were performed in the thermomixer at  $45\text{ }^{\circ}\text{C}$  for 1.5 h at 600 rpm. Subsequent purification was performed using a QIAquick PCR Purification Kit from Qiagen.

**PCR Amplification.** PCR reactions to generate dsDNA libraries were prepared using a modified Taq DNA Polymerase (LongAmp Taq 2 $\times$  Master Mix from New England Biolabs) according to the manufacturer's instructions, with the following profile: 30 s at  $94\text{ }^{\circ}\text{C}$ ; 35 [20 s at  $94\text{ }^{\circ}\text{C}$ , 30 s at  $57\text{ }^{\circ}\text{C}$ , 1–2 min at  $65\text{ }^{\circ}\text{C}$ ]; final extension 10 min at  $65\text{ }^{\circ}\text{C}$ . Individual primer concentrations (FP1–3 and RP1–6 respectively; see Table S1) for each experiment were 0.4  $\mu\text{M}$ , and 2.6 to 8.2 ng clicked-cDNA was used as the template for each 25  $\mu\text{L}$  reaction. Annealing temperatures were calculated with respect to the specific primer pair using the OligoAnalyzer Tool from Integrated DNA Technologies. Elongation times were calculated according to the anticipated PCR product length using the formula  $t(\text{s}) = \text{fragment size (kB)} \times 60$ . PCR products were imaged via agarose gel electrophoresis in Tris-acetate-EDTA buffer and stained using an aqueous solution of approximately 0.2  $\mu\text{g}/\text{mL}$  ethidium bromide. PCR products were further purified using the QIAquick PCR Purification Kit from Qiagen.

**ONT Sequencing.** Amplified PCR products derived from clicked-cDNAs were converted into sequencing libraries using the Ligation Sequencing Kit (SQK-LSK109), according to the Genomic DNA by Ligation protocol and subsequently sequenced on a MinION flow cell (R10.3) with a MinION sequencing device (Oxford Nanopore Technologies). For individual samples, 1  $\mu\text{g}$  of PCR-amplified cDNA was carried through to the first step of the kit protocol.

For standard mRNA sequencing without click chemistry (eGFP mRNA: see mRNA Synthesis; Jurkat Total RNA: purchased from Thermo Fischer Scientific) library preparation

was performed according to the cDNA-PCR-Sequencing (SQK-PCS109) kit instructions from Oxford Nanopore Technologies. The supplied protocol employs strand switching (TS) with kit-supplied oligonucleotides, as with the Smart-Seq2 protocol. The library was subsequently sequenced on a MinION flow cell (R10.3) with a MinION sequencing device (Oxford Nanopore Technologies).

The individual steps of the commercially available kits were performed as directed, with the exception of sample purification, where we instead employed CleanNGS beads from CleanNA instead of the recommended product Agencourt AMPure XP beads from Beckman Coulter.

**Bioinformatic Analysis.** Jurkat cDNA reads derived from the TS library on the MinION device were cleaned for adapter contamination with Porechop (<https://github.com/rrwick/Porechop>) with default parameters. The click-library was trimmed with Porechop using the custom adapter sequence ACGCTCTTCCGATCTAC/GTAGATCGGAAGAGCGT which can be found on both cDNA ends after PCR amplification with primer FP1. Mapping of reads against the *Homo sapiens* GRCh38 DNA primary assembly release 105<sup>31</sup> was performed with Minimap2 (<https://github.com/lh3/minimap2>)<sup>32</sup> using the parameter set “-ax splice”. The eGFP amplicon reads were trimmed with Porechop for the custom sequence GACGCTCTTCCGATCT/AGATCGGAAGAGCGTC, which is present on both ends of each fragment. For mapping of the eGFP amplicon (see [Figure S7](#)) with minimap2, the parameter set “-ax map-ont” was used.

Data set statistics were calculated and visualized with NanoComp.<sup>33</sup> Mapping graphics were generated with IGV 2.12.2<sup>34</sup> and assembled with Adobe Illustrator CS6 and Affinity Designer.

Transcript abundance was estimated by generating pseudoalignments against the Ensemble GRCh38 cDNA release 105<sup>31</sup> with kallisto<sup>24</sup> using read-length statistics from the NanoComp output. Gene-level identification was performed in R version 4.2.1 with the *featureCounts* function of the Rsubread package<sup>25</sup> based on the reads mapped with minimap2. Venn diagrams were generated from counts with a cutoff of >10 cpm or tpm respectively in R using the packages limma<sup>35</sup> and VennDiagram<sup>36</sup> and adapted with Adobe Illustrator CS6 and Affinity Designer.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.bioconjchem.2c00353>.

Oligonucleotide sequences, agarose gel images, electropherogram data, map of the plasmid, sequences of the mRNA model system (eGFP), and additional sequencing data analysis (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

Thomas Frischmuth – baseclick GmbH, 82061 Neuried (Munich), Germany; [orcid.org/0000-0001-8355-6458](https://orcid.org/0000-0001-8355-6458); Email: [t.frischmuth@baseclick.eu](mailto:t.frischmuth@baseclick.eu)

### Authors

Eva S. Schönegger – Ludwig-Maximilians-Universität München, Institute for Chemical Epigenetics Munich, 81377 Munich, Germany; [orcid.org/0000-0003-1860-1739](https://orcid.org/0000-0003-1860-1739)

Antony Crisp – baseclick GmbH, 82061 Neuried (Munich), Germany; [orcid.org/0000-0001-7173-4376](https://orcid.org/0000-0001-7173-4376)

Markus Müller – Ludwig-Maximilians-Universität München, Institute for Chemical Epigenetics Munich, 81377 Munich, Germany; [orcid.org/0000-0002-3579-3317](https://orcid.org/0000-0002-3579-3317)

Jessica Fertl – baseclick GmbH, 82061 Neuried (Munich), Germany

Sascha Serdjukow – baseclick GmbH, 82061 Neuried (Munich), Germany

Stefano Croce – baseclick GmbH, 82061 Neuried (Munich), Germany

Michael Kollaschinski – baseclick GmbH, 82061 Neuried (Munich), Germany

Thomas Carell – Ludwig-Maximilians-Universität München, Institute for Chemical Epigenetics Munich, 81377 Munich, Germany; [orcid.org/0000-0001-7898-2831](https://orcid.org/0000-0001-7898-2831)

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.bioconjchem.2c00353>

## Author Contributions

#E.S.S. and A.C. contributed equally to this work.

## Funding

This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 861381 (NATURE-ETN) and 642023 (ClickGene) and by baseclick GmbH.

## Notes

The authors declare the following competing financial interest(s): We declare that Thomas Frischmuth and Thomas Carell are shareholders of baseclick GmbH.

## ■ ACKNOWLEDGMENTS

A.C. and E.S.S. gratefully acknowledge Dr. Pascal Giehr for his thoughtful advice in the preparation of ONT sequencing libraries. Álvaro Cuadros-Inostroza is thanked for conducting preliminary BLAST searches. Clara Groot Crego is thanked for her thoughtful advice concerning RNA-seq data analyses.

## ■ REFERENCES

- Green, M. R. Pre-mRNA Splicing. *Annu. Rev. Genet.* **1986**, *20*, 671–708.
- Kelemen, O.; Convertini, P.; Zhang, Z.; Wen, Y.; Shen, M.; Falaleeva, M.; Stamm, S. Function of Alternative Splicing. *Gene* **2013**, *514* (1), 1–30.
- Tian, Y.; Wen, H.; Qi, X.; Zhang, X.; Liu, S.; Li, B.; Sun, Y.; Li, J.; He, F.; Yang, W.; Li, Y. Characterization of Full-Length Transcriptome Sequences and Splice Variants of Lateolabrax Maculatus by Single-Molecule Long-Read Sequencing and Their Involvement in Salinity Regulation. *Front. Genet.* **2019**, *10*, 1226 DOI: [10.3389/fgene.2019.01126](https://doi.org/10.3389/fgene.2019.01126).
- Harrow, J.; Frankish, A.; Gonzalez, J. M.; Tapanari, E.; Diekhans, M.; Kokocinski, F.; Aken, B. L.; Barrell, D.; Zadissa, A.; Searle, S.; et al. GENCODE: The Reference Human Genome Annotation for the ENCODE Project. *Genome Res.* **2012**, *22* (9), 1760–1774.
- Oikonomopoulos, S.; Bayega, A.; Fahiminiya, S.; Djambazian, H.; Berube, P.; Ragoussis, J. Methodologies for Transcript Profiling Using Long-Read Technologies. *Front. Genet.* **2020**, *11*, 606 DOI: [10.3389/fgene.2020.00606](https://doi.org/10.3389/fgene.2020.00606).
- Picelli, S.; Björklund, Å. K.; Faridani, O. R.; Sagasser, S.; Winberg, G.; Sandberg, R. Smart-Seq2 for Sensitive Full-Length Transcriptome Profiling in Single Cells. *Nat. Methods* **2013**, *10* (11), 1096–1100.
- Ramsköld, D.; Luo, S.; Wang, Y. C.; Li, R.; Deng, Q.; Faridani, O. R.; Daniels, G. A.; Khrebtkova, I.; Loring, J. F.; Laurent, L. C.;

Schroth, G. P.; Sandberg, R. Full-Length mRNA-Seq from Single-Cell Levels of RNA and Individual Circulating Tumor Cells. *Nat. Biotechnol.* **2012**, *30* (8), 777–782.

(8) Picelli, S.; Faridani, O. R.; Björklund, Å. K.; Winberg, G.; Sagasser, S.; Sandberg, R. Full-Length RNA-Seq from Single Cells Using Smart-Seq2. *Nat. Protoc.* **2014**, *9* (1), 171–181.

(9) Zhu, Y. Y.; Machleder, E. M.; Chenchik, A.; Li, R.; Siebert, P. D. Reverse Transcriptase Template Switching: A SMART Approach for Full-Length cDNA Library Construction. *Biotechniques* **2001**, *30* (4), 892–897.

(10) Saliba, A. E.; Westermann, A. J.; Gorski, S. A.; Vogel, J. Single-Cell RNA-Seq: Advances and Future Challenges. *Nucleic Acids Res.* **2014**, *42*, 8845–8860.

(11) Wu, A. R.; Wang, J.; Streets, A. M.; Huang, Y. Single-Cell Transcriptional Analysis. *Annu. Rev. Anal. Chem.* **2017**, *10*, 439–462.

(12) Wulf, M. G.; Maguire, S.; Humbert, P.; Dai, N.; Bei, Y.; Nichols, N. M.; Corrêa, I. R.; Guan, S. Non-Templated Addition and Template Switching by Moloney Murine Leukemia Virus (MMLV)-Based Reverse Transcriptases Co-Occur and Compete with Each Other. *J. Biol. Chem.* **2019**, *294* (48), 18220–18231.

(13) Cocquet, J.; Chong, A.; Zhang, G.; Veitia, R. A. Reverse Transcriptase Template Switching and False Alternative Transcripts. *Genomics* **2006**, *88* (1), 127–131.

(14) Jaworski, E.; Routh, A. ClickSeq: Replacing Fragmentation and Enzymatic Ligation with Click-Chemistry to Prevent Sequence Chimeras. *Methods Mol. Biol.* **2018**, *1712*, 71–85.

(15) Routh, A.; Head, S. R.; Ordoukhanian, P.; Johnson, J. E. ClickSeq: Fragmentation-Free Next-Generation Sequencing via Click Ligation of Adaptors to Stochastically Terminated 3'-Azido cDNAs. *J. Mol. Biol.* **2015**, *427* (16), 2610–2616.

(16) Fantoni, N. Z.; El-Sagheer, A. H.; Brown, T. A Hitchhiker's Guide to Click-Chemistry with Nucleic Acids. *Chem. Rev.* **2021**, *121* (12), 7122–7154.

(17) El-Sagheer, A. H.; Brown, T. Synthesis and Polymerase Chain Reaction Amplification of DNA Strands Containing an Unnatural Triazole Linkage. *J. Am. Chem. Soc.* **2009**, *131* (11), 3958–3964.

(18) Croce, S.; Serdjukow, S.; Carell, T.; Frischmuth, T. Chemoenzymatic Preparation of Functional Click-Labeled Messenger RNA. *ChemBioChem.* **2020**, *21* (11), 1641–1646.

(19) Kollaschinski, M.; Sobotta, J.; Schalk, A.; Frischmuth, T.; Graf, B.; Serdjukow, S. Efficient DNA Click Reaction Replaces Enzymatic Ligation. *Bioconjugate Chem.* **2020**, *31* (3), 507–512.

(20) Vardi, O.; Shamir, I.; Javasky, E.; Goren, A.; Simon, I. Biases in the SMART-DNA Library Preparation Method Associated with Genomic Poly DA/DT Sequences. *PLoS One* **2017**, *12* (2), e0172769.

(21) Elrod, N. D.; Jaworski, E. A.; Ji, P.; Wagner, E. J.; Routh, A. Development of Poly(A)-ClickSeq as a Tool Enabling Simultaneous Genome-Wide Poly(A)-Site Identification and Differential Expression Analysis. *Methods* **2019**, *155*, 20–29.

(22) Routh, A.; Ji, P.; Jaworski, E.; Xia, Z.; Li, W.; Wagner, E. J. Poly(A)-ClickSeq: Click-Chemistry for next-Generation 3'-End Sequencing without RNA Enrichment or Fragmentation. *Nucleic Acids Res.* **2017**, *45* (12), e112.

(23) El-Sagheer, A. H.; Sanzone, A. P.; Gao, R.; Tavassoli, A.; Brown, T. Biocompatible Artificial DNA Linker That Is Read through by DNA Polymerases and Is Functional in *Escherichia Coli*. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (28), 11338–11343.

(24) Bray, N. L.; Pimentel, H.; Melsted, P.; Pachter, L. Near-Optimal Probabilistic RNA-Seq Quantification. *Nat. Biotechnol.* **2016**, *34* (5), 525.

(25) Liao, Y.; Smyth, G. K.; Shi, W. The R Package Rsubread Is Easier, Faster, Cheaper and Better for Alignment and Quantification of RNA Sequencing Reads. *Nucleic Acids Res.* **2019**, *47* (8), e47.

(26) Li, X.; Xiong, X.; Wang, K.; Wang, L.; Shu, X.; Ma, S.; Yi, C. Transcriptome-Wide Mapping Reveals Reversible and Dynamic N1-Methyladenosine Methylome. *Nat. Chem. Biol.* **2016**, *12* (5), 311.

(27) Schwartz, S.; Bernstein, D. A.; Mumbach, M. R.; Jovanovic, M.; Herbst, R. H.; León-Ricardo, B. X.; Engreitz, J. M.; Guttman, M.;

Satija, R.; Lander, E. S.; Fink, G.; Regev, A. Transcriptome-Wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of ncRNA and mRNA. *Cell* **2014**, *159* (1), 148.

(28) Carlile, T. M.; Rojas-Duran, M. F.; Zinshteyn, B.; Shin, H.; Bartoli, K. M.; Gilbert, W. v. Pseudouridine Profiling Reveals Regulated mRNA Pseudouridylation in Yeast and Human Cells. *Nature* **2014**, *515* (7525), 143.

(29) Lucks, J. B.; Mortimer, S. A.; Trapnell, C.; Luo, S.; Aviran, S.; Schroth, G. P.; Pachter, L.; Doudna, J. A.; Arkin, A. P. Multiplexed RNA Structure Characterization with Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension Sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (27), 11063.

(30) Jensen, M. K.; Elrod, N. D.; Yalamanchili, H. K.; Ji, P.; Lin, A.; Liu, Z.; Wagner, E. J. Application and Design Considerations for 3'-End Sequencing Using Click-Chemistry. *Methods Enzymol.* **2021**, *655*, 1–23, DOI: 10.1016/bs.mie.2021.03.012.

(31) Cunningham, F.; Allen, J. E.; Allen, J.; Alvarez-Jarreta, J.; Amode, M. R.; Armean, I. M.; Austine-Orimoloye, O.; Azov, A. G.; Barnes, I.; Bennett, R. Ensembl 2022. *Nucleic Acids Res.* **2022**, *50* (D1), D988.

(32) Li, H. New Strategies to Improve Minimap2 Alignment Accuracy. *Bioinformatics* **2021**, *37* (23), 4572.

(33) de Coster, W.; D'Hert, S.; Schultz, D. T.; Cruts, M.; van Broeckhoven, C. NanoPack: Visualizing and Processing Long-Read Sequencing Data. *Bioinformatics* **2018**, *34* (15), 2666.

(34) Robinson, J. T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E. S.; Getz, G.; Mesirov, J. P. Integrative Genomics Viewer. *Nat. Biotechnol.* **2011**, *29*, 24.

(35) Ritchie, M. E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C. W.; Shi, W.; Smyth, G. K. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* **2015**, *43* (7), e47.

(36) Chen, H. *VennDiagram: Generate High-Resolution Venn and Euler Plots*. <https://CRAN.R-project.org/package=VennDiagram> (accessed 2022-06-15).

## ☐ Recommended by ACS

### Method for Identifying Sequence Motifs in Pre-miRNAs for Small-Molecule Binding

Yusuke Takashima, Kazuhiko Nakatani, *et al.*

SEPTEMBER 23, 2022  
ACS CHEMICAL BIOLOGY

READ 

### Supernatural: Artificial Nucleobases and Backbones to Program Hybridization-Based Assemblies and Circuits

Miguel López-Tena, Nicolas Winssinger, *et al.*

JULY 20, 2022  
BIOCONJUGATE CHEMISTRY

READ 

### Short Oligonucleotides Facilitate Co-transcriptional Labeling of RNA at Specific Positions

Siyu Wang, Yu Liu, *et al.*

MARCH 16, 2022  
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

READ 

### DNA-Templated Reactions with High Catalytic Efficiency Achieved by a Loss-of-Affinity Principle

Dino Gluhacevic von Krüchten, Oliver Seitz, *et al.*

JUNE 13, 2022  
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

READ 

Get More Suggestions >

### 3.2. Second Publication

**TITLE:** Orthogonal End Labelling of Oligonucleotides through Dual Incorporation of Click-Reactive NTP Analogues.

**AUTHORS:** Eva S. Schönegger<sup>+</sup>, Antony Crisp<sup>+</sup>, Marco Radukic, Jonas Burmester, Thomas Frischmuth, and Thomas Carell\*

*+ authors contributed equally (shared 1<sup>st</sup> authorship). \* corresponding author.*

**JOURNAL, YEAR, VOLUME:**

ChemBioChem, **2023**, 25 (1), e202300701 (<https://doi.org/10.1002/cbic.202300701>)

**SUMMARY:**

Modifying nucleic acid structures post-synthetically with a clickable functionality enables a broad range of applications and advanced properties, such as immune system evasion, enhanced stability, fluorescent labelling, chemical 5'-RNA capping, and the development of functional aptamers.

While some chemoenzymatic methods exist for labelling the 3'-end with azido and alkynyl groups, equivalent strategies for the 5'-end are unknown, since the current methods are either inefficient, complex, or involve harsh chemical conditions. To overcome these limitations, we introduce a modular and straightforward technology for sequentially modifying DNA and RNA strands at both ends with click-reactive groups. Our approach utilizes  $\gamma$ -modified ATP analogs, which enable the T4PNK-catalyzed 5'-modification of oligonucleotides. This process is compatible with TdT-catalyzed 3'-elongation using 3'-azido-2',3'-ddGTP. Furthermore, we showed that our method can be applied for both, oligo-oligo ligations and single-stranded DNA circularization. We envision that these approaches will pave the way for the synthesis of highly functionalized oligonucleotides, thereby enhancing the therapeutic and diagnostic potential of oligonucleotides, such as in the context of next-generation sequencing.

### **PERSONAL CONTRIBUTIONS:**

- Conceptualization of the project outline.
- Synthesis of  $\gamma$ -phosphate modified ATP analogue and SPS of ORN1 and ORN2.
- Established and improved T4 PNK catalyzed 5'-end labelling reactions .
- Performed all CuAAC and SPAAC click reactions of 5'-click functionalized oligonucleotides.
- Lead contribution to data curation, writing and preparation of the original manuscript and supporting information together with A. Crisp.
- Edited and revised the manuscript together with all authors.

### **PERMISION AND LICENSING**

Material from: Schönegger *et al.*, Orthogonal End Labelling of Oligonucleotides through Dual Incorporation of Click-Reactive NTP Analogues, <https://doi.org/10.1002/cbic.202300701> ChemBioChem published 2023 by Wiley-VCH GmbH. This article was published under the Creative Commons license CC BY-NC 4.0.

# Orthogonal End Labelling of Oligonucleotides through Dual Incorporation of Click-Reactive NTP Analogues

Eva S. Schönegger<sup>+, [a, b]</sup> Antony Crisp<sup>+, [b]</sup> Marco Radukic,<sup>[c]</sup> Jonas Burmester,<sup>[b]</sup> Thomas Frischmuth,<sup>\*, [b]</sup> and Thomas Carell<sup>\*, [a]</sup>

Post-synthetic modification of nucleic acid structures with clickable functionality is a versatile tool that facilitates many emerging applications, including immune evasion, enhancements in stability, fluorescent labelling, chemical 5'-RNA-capping and the development of functional aptamers. While certain chemoenzymatic approaches for 3'-azido and alkynyl labelling are known, equivalent 5'-strategies are either inefficient, complex, or require harsh chemical conditions. Here, we present a modular and facile technology to consecutively modify DNA and RNA strands at both ends with click-modifiable

functional groups. Our approach using  $\gamma$ -modified ATP analogues facilitates T4 PNK-catalysed 5'-modification of oligonucleotides, a process that is compatible with TdT-catalysed 3'-elongation using 3'-azido-2',3'-ddGTP. Finally, we demonstrate that our approach is suitable for both oligo-oligo ligations, as well as ssDNA circularization. We anticipate that such approaches will pave the way for the synthesis of highly functionalised oligonucleotides, improving the therapeutic and diagnostic applicability of oligonucleotides such as in the realm of next-generation sequencing.

## Introduction

Over the past decades, site-specific bioconjugation of nucleic acids has allowed the burgeoning field of nucleic acid technology to occupy previously unimaginable niches, such as cellular tracking, encoding of small molecule libraries, and the synthesis of targeted oligonucleotide siRNA therapeutics.<sup>[1]</sup> These and other innovations rely entirely upon our ability to introduce specific chemical modifications into nucleic acid molecules, either during their synthesis, or post-synthetically. When it comes to post-synthetic modification of the 3'-ends of DNA and RNA molecules, an assortment of chemical and enzymatic options are available, with the most notable example being polymerase-mediated incorporation of a modified nucleoside triphosphate. 5'-end labelling, meanwhile, is typically limited by stochastic techniques such as nonspecific chemical

crosslinking.<sup>[2]</sup> Motivated by the absence of suitable methods, we set out to develop a general click chemistry-based chemoenzymatic procedure to modify the 5'-end of single-stranded nucleic acids (Figure 1). A key requirement was for us that the new method is compatible with existing nucleotidyltransferase mediated 3'-ligation reactions, such as those pioneered by the Kukhanova laboratory (Figure 1).<sup>[3–7]</sup> Such a method would allow enhanced detectability of oligonucleotides,<sup>[8]</sup> and could pave the way for dual 3'- and 5'-adapter ligations as required for next-generation sequencing library preparation.<sup>[9]</sup> In addition, such a method could enhance antibody and payload ligations to oligonucleotides,<sup>[10]</sup> post-synthetic 5'-capping, chemical gene synthesis,<sup>[11]</sup> and, as investigated here, nucleic acid circularisation.<sup>[9]</sup> The click-chemistry based method would also complement recent 5'-labelling approaches by the Rentmeister laboratory (Figure 1) that are based on promiscuous

[a] E. S. Schönegger,<sup>+</sup> Prof. Dr. T. Carell  
Department Chemie  
Ludwig-Maximilians-Universität München  
Butenandtstr. 5–13, 81377 Munich (Germany)  
E-mail: Thomas.carell@lmu.de

[b] E. S. Schönegger,<sup>+</sup> Dr. A. Crisp,<sup>+</sup> J. Burmester, Dr. T. Frischmuth  
baseclick GmbH  
Floriansbogen 2–4, 82061 Neuried (Germany)  
E-mail: t.frischmuth@baseclick.eu

[c] Dr. M. Radukic  
Technische Fakultät  
Universität Bielefeld  
Universitätsstraße 25, 33615 Bielefeld (Germany)

[<sup>+</sup>] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cbic.202300701>

© 2023 The Authors. ChemBioChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

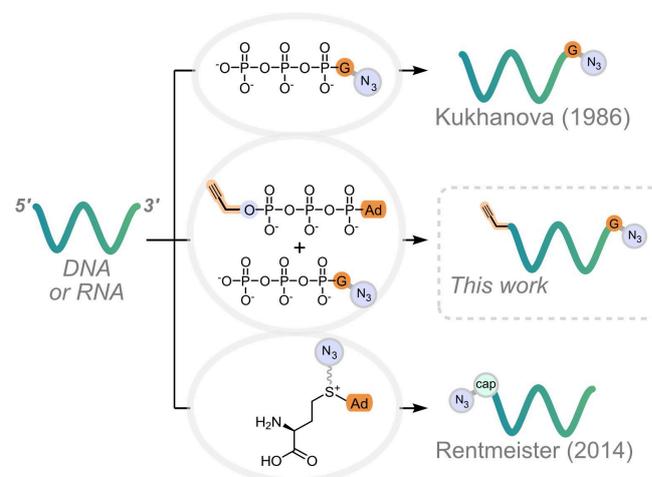


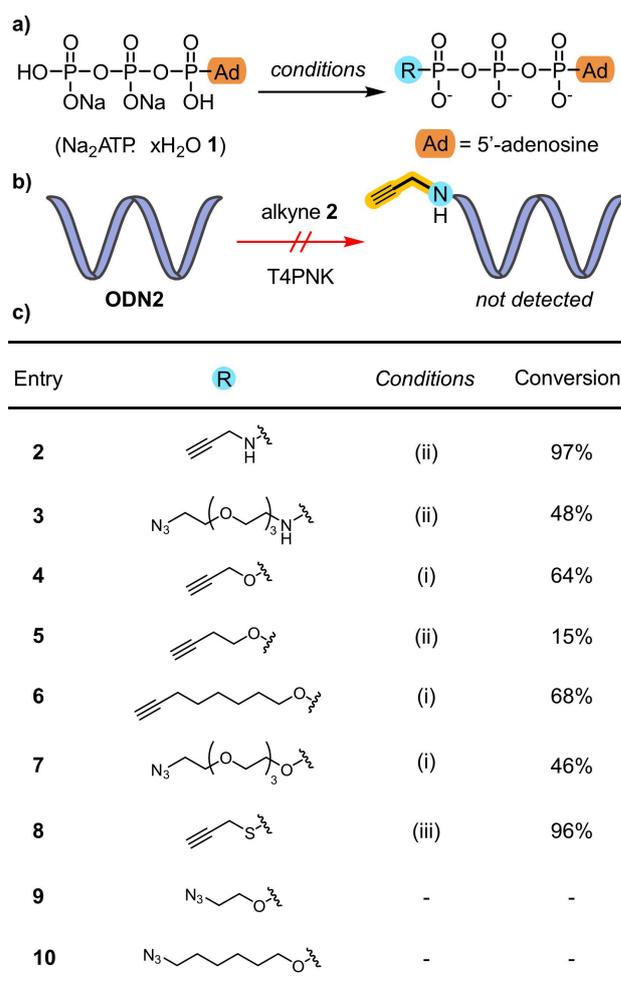
Figure 1. Schematic representation of chemoenzymatic oligonucleotide end-labelling methods that are currently in use.

methyltransferase (Ecm1).<sup>[8]</sup> Furthermore, our approach has the advantage that it involves only one synthetic step per label, and minimal chemical alteration to the DNA or RNA structure (Figure 1). We anticipated that this could be accomplished via the T4 PNK-catalysed transfer of an alkyne- or azido-modified  $\gamma$ -phosphate group from adenosine triphosphate (ATP)<sup>[12–14]</sup> to 5'-hydroxy groups of nucleotides and deoxyribonucleotides – a process that is inherently compatible with TdT-catalysed 3'-elongation reactions (Figure 1).

## Results and Discussion

Motivated by the great potential of  $\gamma$ -phosphate modified ATP analogues for protein kinase catalysed reactions advertised by Lee and co-workers,<sup>[15]</sup> we prepared a small collection of suitably  $\gamma$ -labelled adenosine 5'-triphosphates, with embedded click- functionality, according to a method previously published by our laboratory. In that study, fluorophores were coupled to  $\gamma$ -modified ATP-analogues by means of Cu<sup>I</sup>-catalysed alkyne-azide cycloaddition (CuAAC) reactions.<sup>[16]</sup> For the synthesis, treatment of adenosine 5'-triphosphate (ATP) disodium salt hydrate **1** (Figure 2) with propargylamine in the presence of EDC-HCl afforded compound **2** in a highly satisfactory yield (97%), as determined by analytical RP-HPLC. In order to prepare **2** for use with the enzyme T4-PNK, an ion-exchange method was applied to achieve efficient desalting (Chelex<sup>®</sup> 100 sodium form), followed by a dialysis for 12 h. We next prepared compound **3** using the same method, which provided **3** in a more modest, albeit practicable yield (48%). Originally, phosphoramidates were synthesised, based on their potential application as T4 PNK substrate published by Anthony and co-workers.<sup>[17]</sup> However, in anticipation that phosphoramidates such as **2** or **3** might be prone to acid-mediated solvolysis,<sup>[18,19]</sup> we further prepared a subset of phosphate monoester-linked alkynes (compounds **3–6**) and azide (compound **7**), which was obtained in up to 68% yield by nucleophilic substitution, involving either the corresponding alkyl halide or *p*-toluenesulfonate (Supporting Information, Section 3.2). In addition, we prepared the organothiophosphate **8** using similar conditions (Supporting Information). Finally, compounds **9** and **10** were purchased for the purpose of further reactivity screening.

Having this library of compounds in hand, we next investigated their applicability for 5'-labelling followed by a CuAAC click reaction.<sup>[20]</sup> Initially, we synthetically prepared the longer 43-mer oligonucleotide **ODN1** and the shorter 17-mer nucleotide **ODN2** (Table S1 in the Supporting Information). For the first experiment, **ODN2** was incubated with alkyne **2** and T4 PNK (37 °C, 50 U). Upon the anticipated completion of the enzymatic reaction after 3 h, the resulting 5'-modified oligodeoxynucleotide was subsequently treated with either CalFluor-647-azide or AzDye<sup>™</sup>-350-azide in the presence of Cu<sup>I</sup> to initiate the click reaction. Contrary to our expectations, however, neither experiment yielded any triazole-containing cycloadduct, as inferred from denaturing polyacrylamide gel electrophoresis (dPAGE) images. Instead, mass-spectrometric (MALDI-TOF) data hinted at  $\gamma$ -phosphoramidyl-hydrolysis, which provided the *m/z*



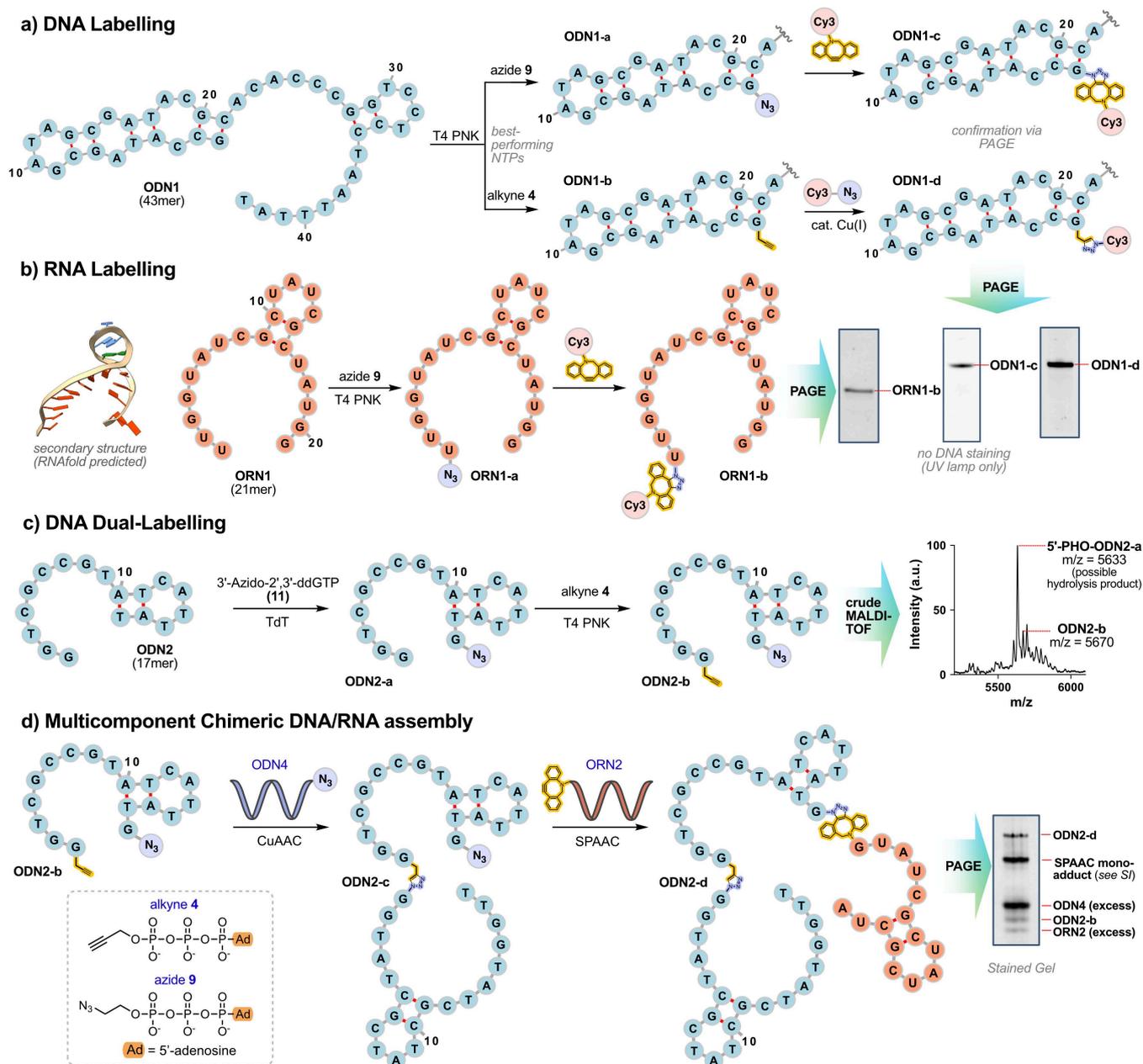
\*conversion monitored by analytical RP-HPLC.

**Figure 2.** a) Synthesis of azide and alkyne  $\gamma$ -labelled nucleoside 5'-triphosphates. b) Unsuccessful first labelling experiment with alkyne **2** on **ODN2**. c) Compound library, reagents and conditions: (i) Step 1: 1, CHELEX-100, [Bu<sub>4</sub>N]Br, Step 2: alkylating tosylate/bromide reagent (Supporting Information), anhydrous DMF, overnight, Ar, RT; (ii) 1, EDC-HCl, alkylating amine, H<sub>2</sub>O, (DMF), 18 h, 25 °C; (iii) ATP $\gamma$ S, DMF/D<sub>2</sub>O, NaHCO<sub>3</sub>, propargyl bromide, RT, overnight.

values for the 5'-phosphorylated oligodeoxynucleotide **ODN2** (5'-PHO-**ODN2**) as the sole reaction product. This result is consistent with previous observations of a rapid hydrolysis of phosphoramidates under mildly acidic conditions.<sup>[18,19]</sup> Similar outcomes were also observed when the azide **3** was employed as the reactive triphosphate (Table S2).

We therefore began to experiment with  $\gamma$ -O-ATP esters such as compounds **9** and **4**, anticipating that these might not suffer the same drawbacks as their  $\gamma$ -N-ATP analogues, and given their reported comparative stabilities at pH values as low as 2.0.<sup>[19]</sup> Gratifyingly, when the same process was conducted using the 43-mer **ODN1** with azide **9** as the substrate, the product **ODN1-a** was indeed obtained in up to 37% yield (Figure 3, Table S3).

When treated with alkyne **4**, the yield of the product **ODN1-b** (**ODN1\_propargyl-PHO**) was higher (57%, Table S3). In order



**Figure 3.** Schematic illustration of several different applications of sequential 3'- and 5'-labelling. a) A DNA 43-mer (ODN1) is labelled at the 5'-end with either an alkyne or an azide, and subsequently click-ligated with a fluorophore (Cy3) for dPAGE visualisation (Figures S1 and S2). b) An RNA 21-mer (ORN1) is labelled at the 5'-end with an azide, and subsequently ligated with Cy3 (Figure S3). c) A DNA 17-mer (ODN2) is labelled with an azide (11) at the 3'-end, and then an alkyne (4) at the 5'-end (Table S5). d) The product from the previous reaction is subsequently treated with two new alkyne- and azido-labelled oligonucleotides, thus affording a new DNA–RNA hybrid product assembled from three oligonucleotide components (for full gel image see Figure S4). For illustrative purposes, the depicted secondary structures were calculated using either RNAComposer<sup>[21,22]</sup> or UNAFold<sup>[23]</sup> using the default parameters, and then visualised using forna<sup>[24]</sup> and ChimeraX<sup>[25]</sup>. ● = DNA, ● = RNA.

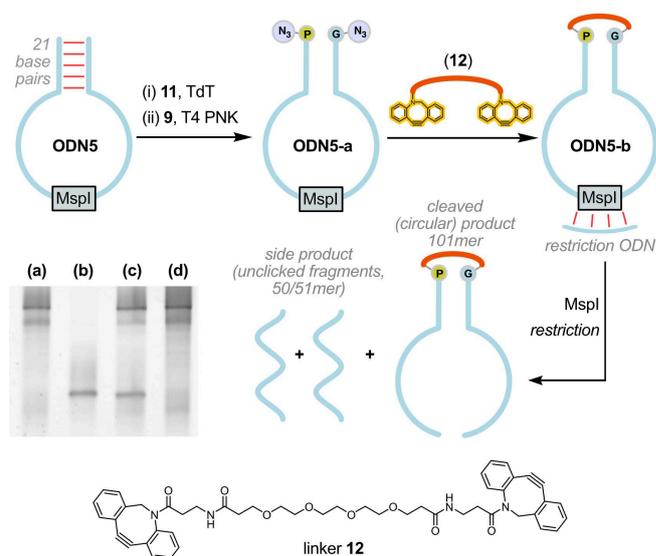
to gain further proof for the successful enzymatic modification and subsequent click reaction, we next carried out click reactions using the fluorophore Cy3 which allows visualisation of the reaction outcome by dPAGE. For the experiment we treated ODN1-a with dibenzocyclooctyne–Cy3 (DBCO–Cy3) and indeed cleanly afforded ODN1-c, as confirmed by dPAGE, both with and without GelRed stain (Figures 3a and S2, confirmed by RP-HPLC, Table S4). Encouraged by this result, we subsequently performed an analogous experiment using ODN1-b in combi-

nation with Cy3-azide, which once again, yielded the dye-adduct ODN1-d in a clean reaction (Figures 3a and S1).

Given that single-stranded ribonucleic acids are known to be natural substrates of T4 PNK, we were curious as to whether a suitable oligoribonucleotide (ORN) could also be labelled. Thus, we next conducted an experiment where we treated the RNA 21-mer ORN1 (Table S1) with azide 9 in the presence of T4 PNK. To our delight, this provided the labelled product, which was visualised via a SPAAC reaction,<sup>[26]</sup> involving either

DBCO–Cy3, or alternatively a 3'-DBCO labelled oligodeoxynucleotide (**ODN3**, Table S1), which produced the expected 21 nucleotide shift upon dPAGE analysis (Figure S3). The yield of the 5'-end labelling and thus also the yield of the subsequent SPAAC is lower compared to the performed experiments with **ODN1** and **ODN2** (Figure S2). This can be explained by well-known observations, that the labelling efficiency by T4 PNK is sequence specific. Thus, factors, such as the identity of the 5'-terminal nucleotide or the ORN folding state can have an impact on the 5'-labelling efficiency.<sup>[27]</sup> A natural extension of this strategy was to assemble a longer hybrid-DNA/RNA oligonucleotide by incrementally combining two DNA and one RNA oligonucleotide. We envisaged that this could be accomplished by site-selectively labelling both the 5'- and 3'-ends of a template oligo with distinct azido and alkynyl functionality followed by two sequential click reactions. To this end, the 17-mer **ODN2** was first treated with 3'-azido-2',3'-ddGTP (**11**) in the presence of catalytic TdT, which gave upon spin column purification, the anticipated 3'-azido labelled oligodeoxynucleotide **ODN2-a**, as confirmed by MALDI-TOF-MS (Table S5). **ODN2-a** was then treated with alkyne **4** in the presence of T4 PNK, to afford **ODN2-b**. In this case, the crude MALDI-TOF showed formation of both the expected products, as well as another peak at *m/z* 5633 (5'-PHO-**ODN2-a**), possibly indicating hydrolysis of the formed alkyne. Anticipating that this might be caused by the ionisation step during MALDI-TOF analysis, we directly continued with the click reaction. Indeed, the dual click reactions of **ODN2-b** together with first the 3'-azido oligodeoxynucleotide **ODN4**, (10× molar excess, Table S1) and second with the 5'-DBCO modified oligoribonucleotide **ORN2** (2× molar excess, Table S1) provided the chimeric DNA/RNA product **ODN2-d**, with only traces of 5'-unligated starting material (visible by dPAGE, Figure S4). This strengthened our hypothesis that the dominant signal in the crude MALDI-TOF corresponded to the ionisation-induced 5'-hydrolysis product. The double chemoenzymatic addition of azide and alkyne functionalities and even the double click reaction had worked as inferred from the fact that the correct oligonucleotide was among the dominant species detected by dPAGE, taken into account, that **ODN4** and **ORN2** were added in excess (Figures 3d and S4).

In a further experiment meant to examine the synthetic utility of our method, we set the goal to achieve circularisation of a medium length (100 nt) oligodeoxynucleotide, **ODN5** (Table S1), which was specifically selected for generating a self-annealing stem loop at the ends. Beyond their many biological roles,<sup>[28]</sup> circular nucleic acids have recently attracted enormous interest due to their asserted exonuclease resistance,<sup>[29]</sup> which makes them highly attractive candidates as novel RNA therapeutics.<sup>[30]</sup> In order to accomplish this task, **ODN5** was initially both 5'- and 3'-azido labelled, according to the procedure described above to afford **ODN5-a** (Figure 4). The purified and labelled oligodeoxynucleotide was then reannealed, and treated with the bifunctional crosslinker **12**, DBCO-PEG4-DBCO, to give **ODN5-b**. Upon completion of the reaction, we proved the circularization by treating the reaction mixture with *MspI*, a restriction enzyme that recognises the sequence CCGG. Remarkably, 10% dPAGE of digested and undigested

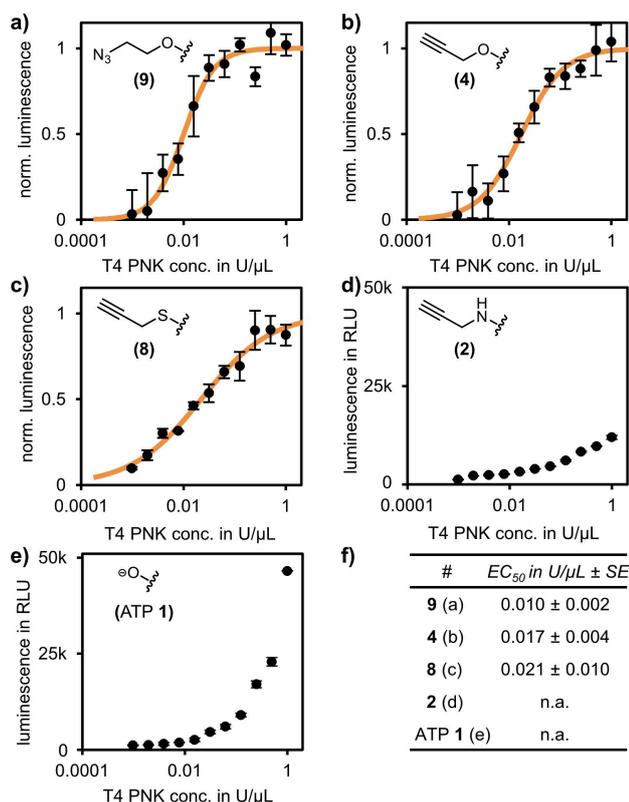


**Figure 4.** Schematic illustration of **ODN5** circularisation and subsequent proof of concept with *MspI*. 10% dPAGE: a) **ODN5**: 100-nt oligonucleotide (Supporting Information), b) **ODN5** treated with *MspI* to give 50-nt oligonucleotide digestion products, c) Crude reaction product containing **ODN5**, **ODN5-a**, and **ODN5-b** treated with *MspI* give mixture of circular 101-nt and 50-/51-nt linear digestion products, d) Crude mixture containing **ODN5**, **ODN5-a**, and **ODN5-b**, without *MspI* digestion. P=PHO (mono-phosphate), G–N<sub>3</sub> = 3'-azido-2',3'-ddGMP.

**ODN5** showed an almost complete disappearance of the 100 nt upper band, whereas the clicked product **ODN5-b** did not, thus indicating a successful circularization process (for full gel image, see Figure S5).

Beyond the  $\gamma$ -modified ATP analogues alkyne **4** and azide **9**, we also screened the other synthesised candidates for their ability to perform 5'-labelling of the 43-mer **ODN1**. The screening revealed that shorter  $\gamma$ -O-linked phosphate esters such as **4** and **9** perform far better as substrates for the T4 PNK catalysed kinase reaction than those with slightly longer linkers. **ODN1** for example was labelled in only 19% and 7% conversions with alkyne **5** and azide **10**, respectively (Table S3). The even longer linkers produced no detectable product. Based on these results, we postulated that the active pocket of the T4 PNK might be sterically constrained, such that larger or more hydrophobic linkers are not suitable to 5'-labelling. To further investigate our hypothesis, we determined binding characteristics of the compounds to T4 PNK.

One indirect but relatively straightforward way to achieve this goal is to measure T4 PNK-mediated hydrolysis of the compounds, enabled by an ADP-dependent luciferase assay (ADP-Glo assay, Figure 5). In this assay, we observed that the azido and alkynyl esters **9** and **4** indeed showed the expected sigmoidal curve of the luminescence signal depended on T4 PNK concentration. The obtained EC<sub>50</sub> values as proxy metrics for their affinities towards T4 PNK were also similar with 0.010 ± 0.002 U  $\mu$ L<sup>-1</sup> for azide **5** and 0.017 ± 0.004 U  $\mu$ L<sup>-1</sup> for alkyne **4**. The N-linked alkyne **2** also showed expectedly no enzyme dependency. Surprisingly, the S-linked alkyne **8** indicated an unexpected luminescence signal intensity and curve shape (Figure 5c), although no ODN labelling was



**Figure 5.** T4 PNK-mediated hydrolysis of ATP analogues under optimised reaction conditions with 1 μM ODN 1, at various enzyme concentrations, determined by an ADP-dependent luciferase assay. The standard T4 PNK concentration is 1 U μL<sup>-1</sup> (the respective rightmost data points). a–c Results are shown as luminescence normalised to the upper and lower asymptote of each fit (orange line, logistic, four-parameter) versus T4 PNK concentration. d, e Results are shown as absolute luminescence signal from the plate reader since no fit was possible. a γ-Azido analogue, entry 9, b γ-propargyl analogue, entry 4, c γ-propargyl-thio analogue, entry 8, d γ-propargyl-imido analogue, entry 2, e ATP (1). Note the logarithmic scale of the x-axes; error bars represent the standard deviation from two technical replicates; RLU: relative fluorescence units. f Table of EC<sub>50</sub> values from the fits; SE: standard error.

detected. Instead, it showed a sigmoidal response curve similar to compounds 9 and 4, with a highly comparable EC<sub>50</sub> value of 0.021 ± 0.010 U μL<sup>-1</sup>. We hypothesise that S-linked alkyne 8 is hydrolysed by T4 PNK independently from the ODN, explaining the large and increasing signal, a finding in support of previous observations.<sup>[17]</sup> Notably, the molecular mechanism of this background hydrolysis is not obvious as it is thought that the γ-phosphate undergoes initial nucleophilic attack, where it is directly transferred to the ODN 5'-OH without first forming a phosphoryl-enzyme intermediate.<sup>[14]</sup> Finally, in order to rationalise the data, we performed rigid receptor docking studies using the ATP analogues described above (Supporting Information). This analysis revealed a high positional variability for the adenosine moieties of all compounds, likely reflecting the known acceptance of T4 PNK towards different NTP substrates.<sup>[12]</sup> Overall, the model provided predictive value, in that relative binding energies appeared to be good predictors of the general possibility of labelling.

## Conclusions

In summary, we have demonstrated the application of a novel chemoenzymatic method for the site-specific 5'-labelling of DNA and RNA molecules with either alkynes or azides. Notably, our method is compatible with a TdT-catalysed 3'-labelling reaction, which opens up the possibility to generate multi-component DNA assemblies. Importantly, our new method allows efficient chemically induced post-synthetic circularisation of DNA as demonstrated here using a ssDNA 100-mer. This is currently not feasible with existing methods, except in highly sequence-specific contexts. Our method offers superior orthogonality to related commercial technologies, such as the 5' EndTag Kit from Vector Laboratories. This is due to the nature of click chemistry, which largely avoids side reactions, such as those that can occur when thiophosphates are employed. These findings thus open up new avenues for the synthesis of novel complex oligonucleotide structures and procedures, including those needed for next-generation sequencing or for the development of new nucleic acid therapeutics.

## Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 861381 (NATURE-ETN). We would like to express our gratitude to Dr. Markus Müller and Dr. Fabio Spada for fruitful scientific discussions. Open Access funding enabled and organized by Projekt DEAL.

## Conflict of Interests

The authors declare no conflict of interests.

## Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

**Keywords:** chemoenzymatic · click chemistry · click-functionalised nucleotides · labelling · nucleic acids

- [1] N. Z. Fantoni, A. H. El-Sagheer, T. Brown, *Chem. Rev.* **2021**, *121*, 7122–7154.
- [2] N. R. Zearfoss, S. P. Ryder, *Methods Mol. Biol.* **2012**, *941*, 181–193.
- [3] S. Croce, S. Serdjukow, T. Carell, T. Frischmuth, *ChemBioChem* **2020**, *21*, 1641–1646.
- [4] E. S. Schönegger, A. Crisp, M. Müller, J. Fertl, S. Serdjukow, S. Croce, M. Kollaschinski, T. Carell, T. Frischmuth, *Bioconjugate Chem.* **2022**, *33*, 1789–1795.
- [5] I. Manuguerra, S. Croce, A. H. El-Sagheer, A. Krissanaprasit, T. Brown, K. V. Gothelf, A. Manetto, *Chem. Commun.* **2018**, *54*, 4529–4532.
- [6] Z. G. Chidgevadze, R. S. Beabealashvili, A. A. Krayevsky, M. K. Kukhanova, *Biochim. Biophys. Acta* **1986**, *868*, 145–152.
- [7] R. S. Beabealashvili, A. V. Scamrov, T. V. Kutateladze, A. M. Mazo, A. A. Krayevsky, M. K. Kukhanova, *Biochim. Biophys. Acta* **1986**, *868*, 136–144.

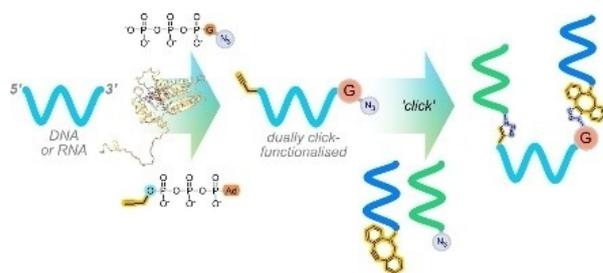
- [8] J. M. Holstein, D. Stummer, A. Rentmeister, *Chem. Sci.* **2015**, *6*, 1362–1369.
- [9] S. Barberán-Soler, J. M. Vo, R. E. Hogans, A. Dallas, B. H. Johnston, S. A. Kazakov, *Genome Biol.* **2018**, *19*, 105.
- [10] I. Dovgan, A. Ehkirch, V. Lehot, I. Kuhn, O. Koniev, S. Kolodych, A. Hentz, M. Ripoll, S. Ursuegui, M. Nothisen, S. Cianférani, A. Wagner, *Sci. Rep.* **2020**, *10*, 7691.
- [11] M. Kukwikila, N. Gale, A. H. El-Sagheer, T. Brown, A. Tavassoli, *Nat. Chem.* **2017**, *9*, 1089–1098.
- [12] C. C. Richardson in *Enzymes*, Academic Press, **1981**, pp. 299–314.
- [13] J. H. Eastberg, J. Pelletier, B. L. Stoddard, *Nucleic Acids Res.* **2004**, *32*, 653–660.
- [14] R. L. Jarvest, G. Lowe, *Biochem. J.* **1981**, *199*, 273–276.
- [15] S. E. Lee, L. M. Elphick, A. A. Anderson, L. Bonnac, E. S. Child, D. J. Mann, V. Gouverneur, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3804–3807.
- [16] S. Serdjukow, F. Kink, B. Steigenberger, M. Tomás-Gamasa, T. Carell, *Chem. Commun.* **2014**, *50*, 1861–1863.
- [17] T. M. Anthony, M. K. H. Pflum, *Bioorg. Med. Chem.* **2018**, *26*, 2331–2336.
- [18] W. S. Wadsworth, *J. Org. Chem.* **1987**, *52*, 1748–1753.
- [19] S. M. Hacker, M. Mex, A. Marx, *J. Org. Chem.* **2012**, *77*, 10450–10454.
- [20] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021.
- [21] M. Popena, M. Szachniuk, M. Antczak, K. J. Purzycka, P. Lukasiak, N. Bartol, J. Blazewicz, R. W. Adamiak, *Nucleic Acids Res.* **2012**, *40*, e112.
- [22] M. Antczak, M. Popena, T. Zok, J. Sarzynska, T. Ratajczak, K. Tomczyk, R. W. Adamiak, M. Szachniuk, *Acta Biochim. Pol.* **2016**, *63*, 737–744.
- [23] N. R. Markham, M. Zuker, *Methods Mol. Biol.* **2008**, *453*, 3–31.
- [24] P. Kerpedjiev, S. Hammer, I. L. Hofacker, *Bioinformatics* **2015**, *31*, 3377–3379.
- [25] E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, T. E. Ferrin, *Protein Sci.* **2021**, *30*, 79–82.
- [26] N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J. Am. Chem. Soc.* **2004**, *126*, 15046–15047.
- [27] V. van Houten, F. Denkers, M. van Dijk, M. van den Brekel, R. Brakenhoff, *Anal. Biochem.* **1998**, *265*, 386–389.
- [28] C.-X. Liu, L.-L. Chen, *Cell* **2022**, *185*, 2016–2034.
- [29] M. S. Xiao, J. E. Wilusz, *Nucleic Acids Res.* **2019**, *47*, 8755–8769.
- [30] J. Li, M. Mohammed-Elsabagh, F. Paczkowski, Y. Li, *ChemBioChem* **2020**, *21*, 1547–1566.

---

Manuscript received: October 13, 2023

Accepted manuscript online: October 20, 2023

Version of record online: ■■, ■■



A new chemoenzymatic method involving dual incorporation of  $\gamma$ - and 3'-modified NTPs allows sequential 5'-

and 3'-end labelling of DNA and RNA molecules.

*E. S. Schönegger, Dr. A. Crisp, Dr. M. Radukic, J. Burmester, Dr. T. Frischmuth\*, Prof. Dr. T. Carell\**

1 – 7

**Orthogonal End Labelling of Oligonucleotides through Dual Incorporation of Click-Reactive NTP Analogues**



## 4. Unpublished Work

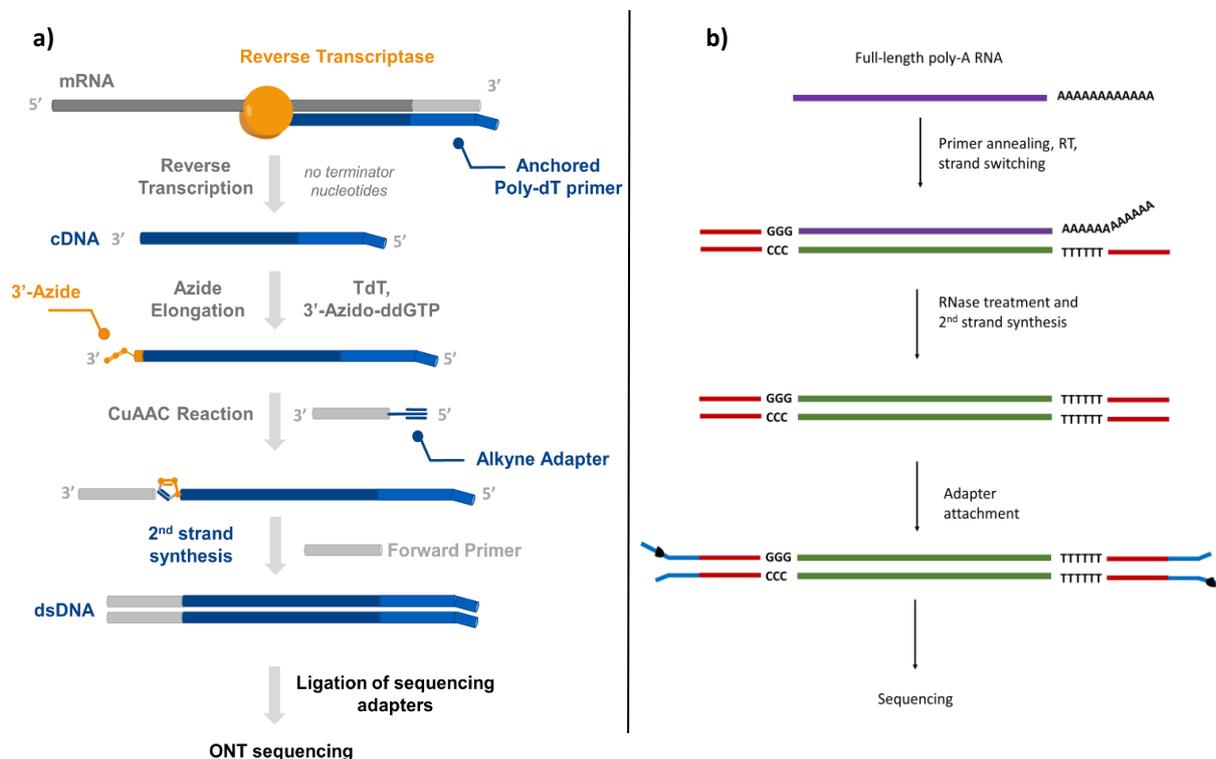
### 4.1. Click Chemistry Based Library Preparation for Direct cDNA Sequencing

#### 4.1.1. Prologue

The results of our first publication dealing with a click chemistry-based library preparation method of full-length reverse transcripts for long-read third generation sequencing,<sup>[78]</sup> prompted us to further investigate possible improvements of this technology. Hence, we got inspired by the direct cDNA sequencing method from ONT (Oxford Nanopore Technologies), which avoids amplification of the generated cDNA and thus removes bias caused by PCR (Figure 14b).<sup>[79]</sup> Previous studies have shown, that when it comes to the amplification of rather complex DNA mixtures, like cDNA libraries, short fragments tend to be amplified more readily compared to longer ones. This causes a false representation of the cDNA library gravitating towards shorter fragments.<sup>[80-81]</sup> Thus, although the direct cDNA sequencing kit (SQK-LSK114) from ONT requires a higher input amount (100 ng Poly(A)+ RNA or 1 µg of total RNA) compared to their cDNA-PCR sequencing kit (SQK-PCS111) (4 ng Poly(A)+RNA or 200 ng total RNA)<sup>[82-83]</sup>, the PCR-free approach is a valuable tool for differential expression analysis of mRNAs and therefore leading to results of superior quality.<sup>[79, 84]</sup> Hence, we not only wanted to avoid template switching (TS) reactions, but also the amplification of the generated cDNA, based on the protocol of ONT.<sup>[85]</sup>

Our revised strategy (Figure 14a) would start, like previously described, with the cDNA synthesis of a mRNA of interest, using an anchored poly-dT primer, followed by the azide elongation with 3'-azido-ddGTP *via* TdT. Next, the 3'-azido-terminated cDNA is clicked to a 5'-hexynyl phosphate moiety containing adapter oligodeoxynucleotide *via* CuAAC. These steps provide a cDNA with known sequences on both ends, enabling subsequent PCR amplification for generating a dsDNA library.<sup>[78]</sup> For the purpose of direct cDNA sequencing, we now aimed to only perform a second strand synthesis in order to provide dsDNA which is necessary for subsequent ONT sequencing.<sup>[86]</sup> Therefore, only one PCR primer, the forward primer, is applied for one PCR cycle to produce the required dsDNA, which should be subsequently ligated to the sequencing adapters, following the ONT protocol.<sup>[82]</sup> With this click chemistry-based direct cDNA sequencing, we aim to get higher-quality results compared to the published method using PCR amplification of the generated cDNA.<sup>[78]</sup> Moreover, this only slightly

modified protocol enables a side by side comparison of direct cDNA and amplified cDNA sequencing, by only altering the numbers of PCR cycle.

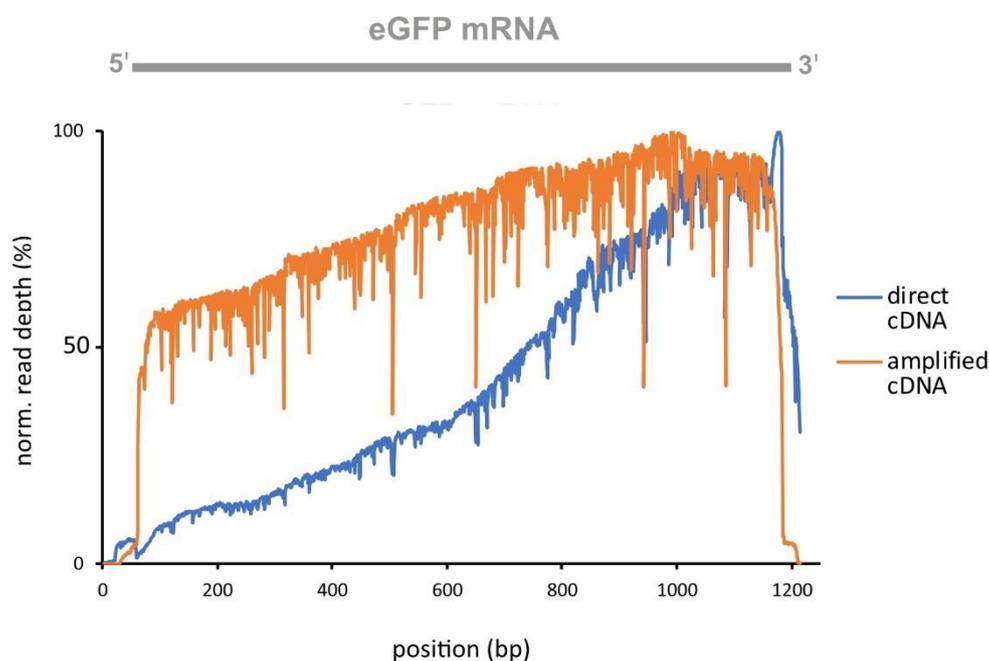


**Figure 14.** **a)** Schematic representation of the direct cDNA library preparation workflow using CuAAC. **b)** Schematic representation of the cDNA-PCR sequencing workflow from ONT.<sup>[82]</sup>

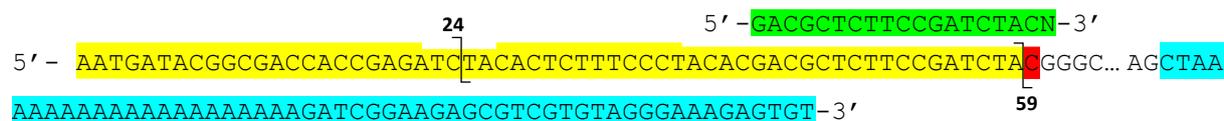
#### 4.1.2. Results and Discussion

For the investigation of the click chemistry based direct cDNA sequencing approach, the same template as for the amplified cDNA sequencing, eGFP mRNA, was used (Schönegger *et al.* Figure S5<sup>[78]</sup>). After reverse transcription with RTP3 (Schönegger *et al.* Table S1<sup>[78]</sup>) followed by the azide elongation with 3'-azido-ddGTP *via* TdT, the 3'-azido-terminated cDNA was clicked to the 5'-hexynyl phosphate moiety containing adapter oligodeoxynucleotide AA1 (Schönegger *et al.* Table S1<sup>[78]</sup>) *via* CuAAC, using the published protocol.<sup>[78]</sup> Afterwards the 2<sup>nd</sup> strand synthesis with FP3 (Schönegger *et al.* Table S1<sup>[78]</sup>) was performed using a modified protocol of the SQK-LSK114 (previously SQK-DCS109) kit from ONT<sup>[82]</sup>, followed by the end repair (end-prep) and adapter ligation step, according to the manufacturers procedure. The adapter ligated dsDNA library was subsequently sequenced for 24 h on a MinION device (ONT) and the processed sequencing data, was mapped against the eGFP amplicon (Figure 16). To

our surprise, the comparison of the coverage depth (read numbers) of the direct cDNA and amplified cDNA generated sequencing data (Figure 15) reveals poorer sequencing quality for the direct cDNA approach. However, the gradually lower coverage depth towards the 5'- end of the eGFP mRNA, might not come from the adjusted protocol itself, but rather from a low quality mRNA sample. Very likely, the transcription of the eGFP plasmid DNA caused low quality 5'- ends of the eGFP mRNA and this subsequently led to uncomplete 3'-end of the corresponding cDNA. The read depth of the direct sequencing data shows an abrupt drop at the position 59 (bp), which would be the position directly after the reverse complementary sequence to the click-adapter (AA1) (Figure 15), the position of triazole backbone. On the other hand, there is also an abrupt drop after 24 nt (position 24 of 5'-3' of second strand cDNA, Figure 16) which cannot be assigned with the existing template. Thus, the abrupt drop at position 59 nt could very likely be an ONT adapter trimming artefact, since the amplified library also drops at this position.<sup>[78]</sup> A recent publication, reporting the successful incorporation of triazole backbones for library preparation towards protein sequencing approaches using nanopores from ONT, support this hypothesis, that the triazole does not cause issues during nanopore sequencing.<sup>[77]</sup> Further, it needs to be mentioned, that for processing the reads of the direct cDNA sequencing approach no click-adapter trimming, as for the amplified cDNA sequencing<sup>[78]</sup>, was performed. Thus, the coverage does not reach zero, but the differences in read processing do not have an impact on the overall interpretation of the results. Overall, the experiment towards direct cDNA sequencing using a click chemistry-based protocol could achieve full coverage, even though the coverage depth is gradually dropping towards the 5'-end.



**Figure 15.** Coverage depth (normalized read numbers) of the eGFP model transcript of two independent runs on a MinION device (blue: direct click chemistry-based cDNA sequencing, orange: PCR amplified click chemistry-based cDNA sequencing).



**Figure 16.** Sequence of double stranded cDNA (5' to 3'; equally to 5'-3' of the original eGFP mRNA) obtained after second-strand synthesis of the click-ligated first-strand cDNA. Structurally relevant sequence segments are color coded in the following way: reverse complementary sequence to the click-adaptor (AA1) in yellow; the forward amplification primer (FP3) in green; complementary C nucleotide of the added 3'-azido-2',3'-ddG in red and the reverse complementary sequence to reverse transcription primer (RT3) in turquoise.

## 4.2. Unpublished Data towards the 2<sup>nd</sup> Publication

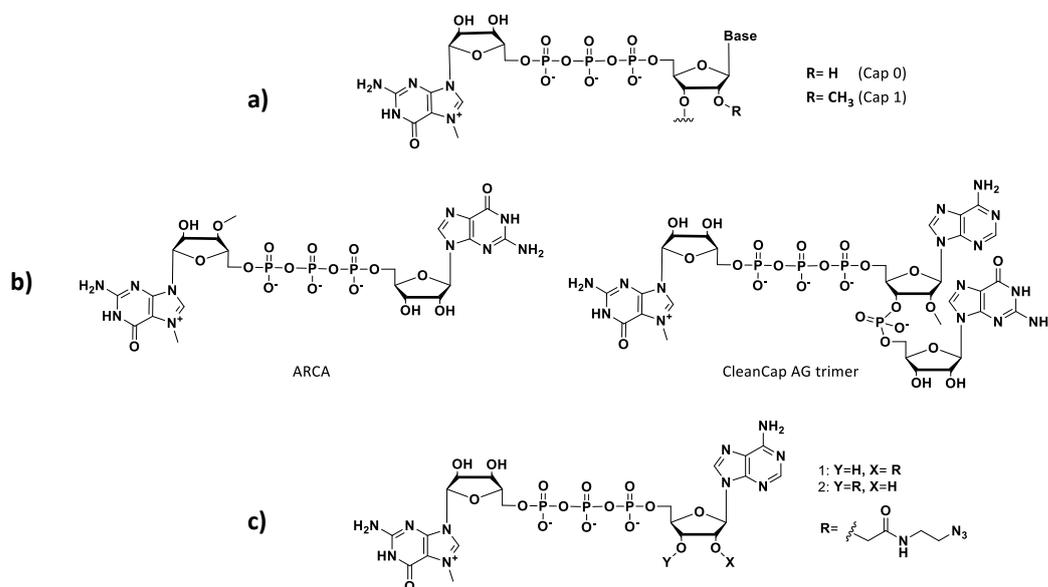
### 4.2.1. Prologue

In the course of study towards the publication of Schönegger *et al.*<sup>[87]</sup>, we were seeking to expand the scope of  $\gamma$ -phosphate modified ATP analogues in order to enable 5' -end labelling of nucleic acid, allowing a click reaction with a molecule of interest. Moreover, we were interested in the preparation of a click-functionalized 5'-cap-analogue, which should represent a straightforward way for post-synthetic 5'-end capping of RNA, simply by clicking it to the 5'-end, which has previously been modified with a clickable moiety *via* T4 PNK catalyzed reaction using a  $\gamma$ -phosphate modified ATP analogue.

The 5'-m<sup>7</sup>G cap, an N<sup>7</sup>-methylated guanosine being linked to the first nucleotide of the mRNA *via* reverse 5' to 5'-triphosphate linkage, displays a conserved modification of eukaryotic mRNA. It plays an essential role in the recruitment of cellular proteins and is responsible for cap- related biological functions, like pre-mRNA processing, nuclear export and cap-dependent protein synthesis and further protects the mRNA from 5' to 3' exonuclease cleavage. The capping takes place co-transcriptionally and in eucaryotic cells the 5' -cap exists either in its cap 0 or cap 1 structure (Figure 17a), where the cap 1, displaying an additional 2'- O methylation, has been attributed to play an important role in evading the cellular immune response.<sup>[88]</sup> Chemical modifications of these cap analogues can influence the translation efficiency, nuclear stability and binding affinity, which, with respect to the recent mRNA vaccine development has emphasized the importance of developing novel tri- and dinucleotide cap analogues.<sup>[89]</sup> In order to avoid that cap analogues are connected with the mRNA strands in the reverse direction, which results in a reduced translation efficiency since the produced reverse-capped mRNA is not recognized by the ribosomes, two chemically modified cap analogues reached high importance. First, the anti-reverse cap analog (ARCA) (m<sup>7</sup>(3'-O-methyl)-GpppG) has been invented by Stepinski and improved by Jemielity and co-workers, in order to be only connected to the 5'-end of mRNA in forward orientation.<sup>[90-92]</sup> This cap 0 analogue has further been improved by the CleanCap<sup>®</sup> cap 1 analogues, which have been developed by TriLink BioTechnologies (Figure 17b).<sup>[91, 93]</sup>

Apart from the development of cap analogues for co-transcriptional 5'- capping, also cap analogues bearing clickable functionalities were synthesized. These hypermethylated azide or

alkyne bearing caps, were attached to oligoribonucleotides *via* Cu(I)catalyzed azide-alkyne cycloaddition, in order to mimic small nuclear (sn) RNAs, bearing a hypermethylated 2,2,7-trimethylguanosine ( $m^{2,2,7}G$ ) cap structure and became of exceptional interest after studies have shown that nuclear delivery of nucleic acids have been enhanced by a synthetic snRNA  $m^{2,2,7}G$ -CAP (Figure 17c).<sup>[94-96]</sup>



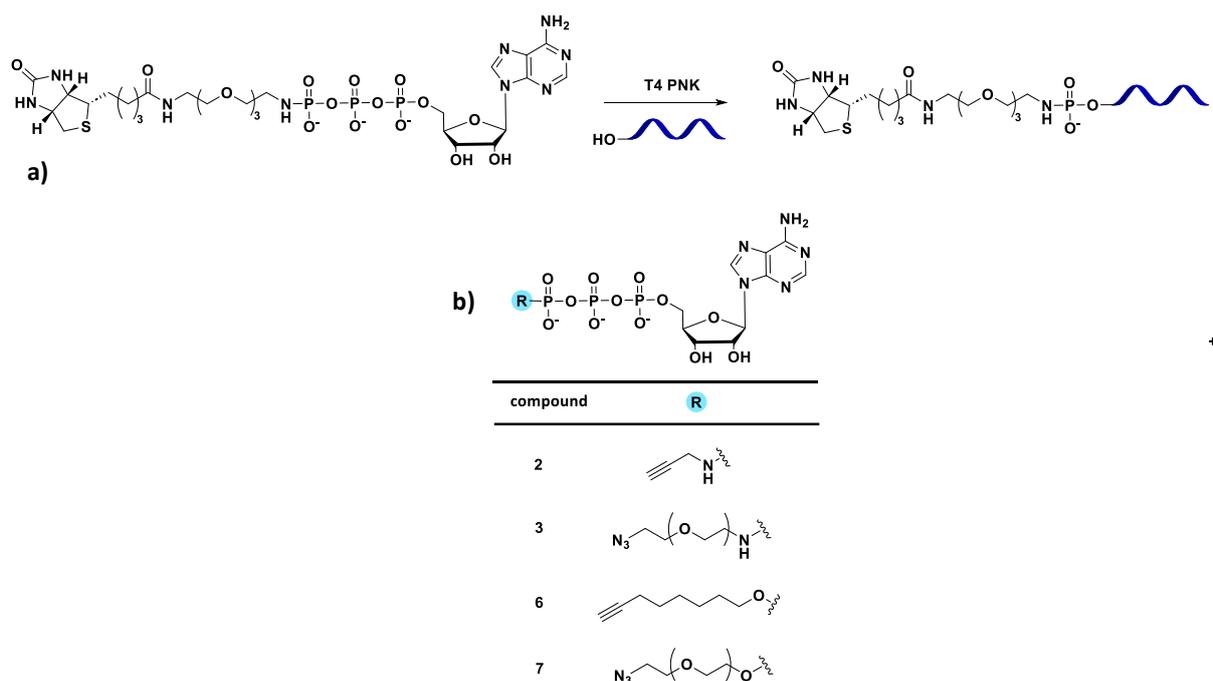
**Figure 17.** a) Structure of cap 0 and cap 1. b) ARCA and CleanCap AG trimer.<sup>[90, 93]</sup> c) Exemplary structure of clickable cap analogues.<sup>[95-96]</sup>

#### 4.2.2. Results and Discussion

##### **$\gamma$ -modified-ATP analogues and their potential application**

Inspired by the work of Anthony and co-workers<sup>[97]</sup>, where a  $\gamma$ -phosphate modified ATP-biotin analogues (Figure 18) was accepted as substrate for the T4 PNK catalyzed 5'-end labelling of ssDNA, we designed our library of click functionalized  $\gamma$ -phosphate ATP analogues based on the chemical entity thereof.<sup>[87]</sup> Based on their  $\gamma$ -phosphate modified ATP-biotin analogue, we first started to synthesize  $\gamma$ -phosphate modified phosphoramidates **2** and **3** of the published work from Schönegger *et al.*<sup>[87]</sup> The synthesis of  $\gamma$ -alkyne-phosphate modified phosphoramidates has previously been published by Serdjukov and co-workers<sup>[73]</sup>, and thus we wanted to apply the synthetic procedure to our phosphoramidates of interest. Especially compound **3** was designed in reference to the  $\gamma$ -modified phosphoramidate introduced by Anthony *et al.*, displaying a similar linker between the  $\gamma$ -phosphate and the molecule of

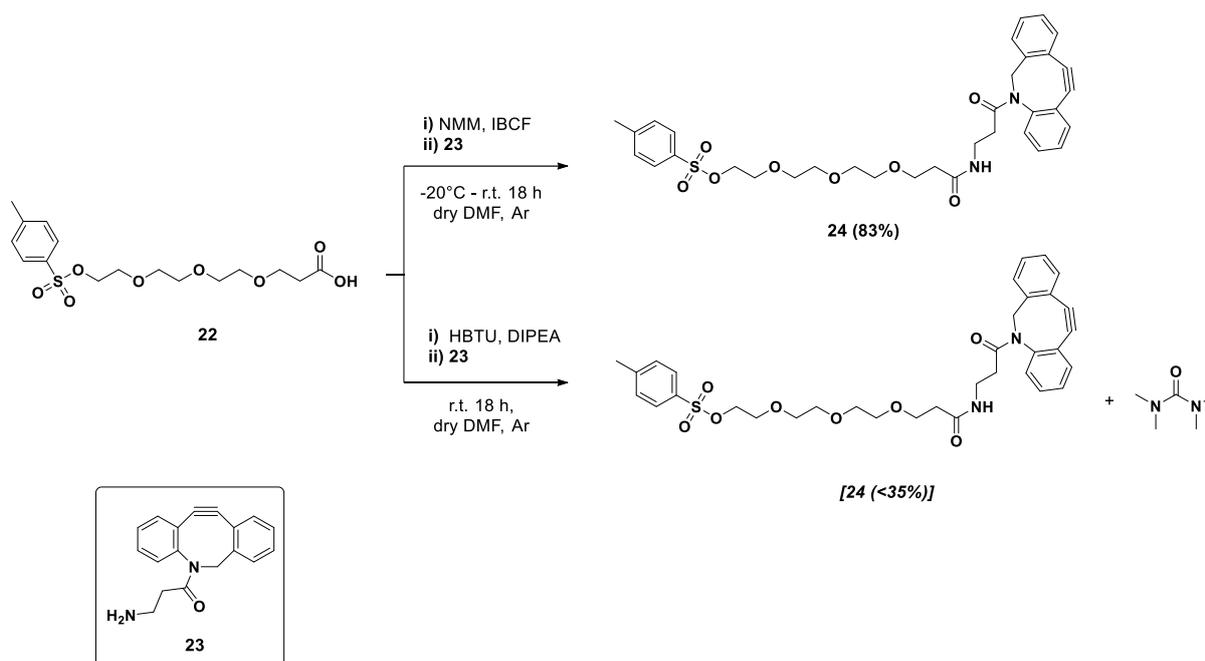
interest in terms of length and structure. In the work of the Pflum laboratory the molecule of interest was biotin, whereas in our approach it should be a clickable functionality, namely  $N_3$  (Figure 18).<sup>[97]</sup> Since we did not achieve successful 5'-end labelling of oligonucleotides with  $\gamma$ -N modified triphosphates due to instability of the N-P bond of the phosphoramidates, we started the preparation of  $\gamma$ -phosphate modified phosphoesters due to the higher stability of the O-P bond.<sup>[87, 98]</sup> Again, we wanted to design the linker in reference to the ATP-biotin analogue from Anthony and co-workers, as we originally thought longer PEG linkers are beneficial for the design of the artificial T4 PNK substrates. Thus, we synthesized compound **7**, with a PEG-3 linker between the  $\gamma$ -phosphate and an azide, enabling subsequent click reactions.<sup>[87]</sup> Since compound **6**, bearing an alkyne functionality, was commercially available, we did not synthesize the alkyne modified analogue of compound **7**, even though the linker of **6** differs in chemical properties and length (alkyl linker and two  $CH_2$  groups shorter than the one of compound **7**).



**Figure 18.** a) Schematic representation of kinase catalyzed biotinylation by Anthony and co-workers.<sup>[97]</sup> b)  $\gamma$ -phosphate modified ATP analogues from Schönegger and co-workers.<sup>[87]</sup>

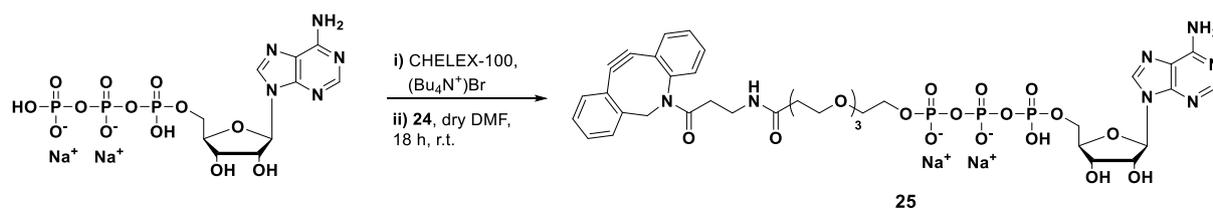
For our initial experiments, however, we first tested **6** and **7** regarding their potential as T4 PNK substrate and synthesize, if necessary, the alkyne bearing analogue of compound **7**. In order to expand the scope of the  $\gamma$ -phosphate modified phosphoesters, we further wanted to synthesize an ATP analogue bearing a DBCO functionality, enabling subsequent straightforward SPAAC with any  $N_3$ -bearing molecule of interest. Successful 5'-end labelling of

single-stranded nucleic acids with a DBCO-ATP analogue *via* T4 PNK catalyzed reaction, could be combined with 3'-end labelling of nucleic acids *via* yeast poly(A) polymerase (yPAP) using 3'-Azido-2',3'-ddATP for RNA or terminal deoxynucleotidyl transferase (TdT) using 3'-azido-2',3'-ddGTP for ssDNA.<sup>[50-51, 58, 99]</sup> The 5'-DBCO and 3'-N<sub>3</sub> functionalized nucleic acids could then undergo SPAAC in order to circularize nucleic acids or to form concatemeric structures, depending on the reaction conditions. For the preparation of the  $\gamma$ -DBCO-PEG4-ATP analogue, we aimed to apply the same synthetic route as described in our publication<sup>[87]</sup>, which was performed in reference to Hacker *et al.*<sup>[98, 100]</sup> Since for this direct alkylation of the  $\gamma$ -phosphate of ATP, either the tosylates or bromides of the alkylating reagents are required<sup>[100]</sup>, compound **24** was synthesized as a precursor for the subsequent S<sub>N</sub>2 reaction. The initial approach for synthesizing **24**, was based on the protocol of Kettenbach *et al.*, where 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) in the presence of *N,N*-diisopropylethylamine (DIPEA) was used as coupling reagent (Scheme 1b).<sup>[101]</sup> The amine- carboxylic acid coupling worked, was however low yielding, with significant amounts of tetramethyl urea impurities. Thus, we tested another coupling reagent, isobutyl chloroformate (IBCF) in combination with 4-methylmorpholine (NMM) as base, which is a standard coupling reagent used for peptide synthesis.<sup>[102]</sup> The reaction afforded **24** with a yield of 83% (Scheme 1a).



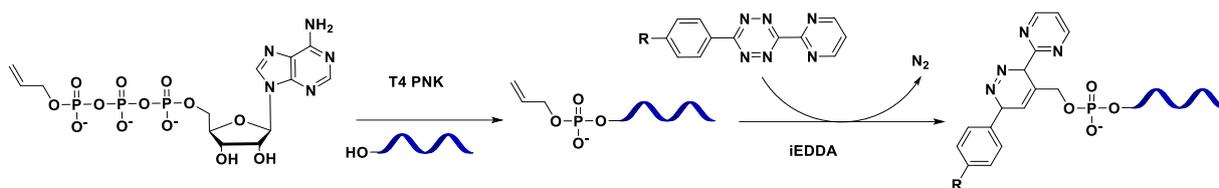
**Scheme 1.** Synthesis of **24**. **a)** Activation of **22** using **i)** IBCF and NMM, followed by the addition of **ii)** **23** at -15°C in dry DMF under Ar atmosphere, allowed to warm to r.t., 18 h, yielding 83%. **b)** activation of **22** using **i)** HBTU and DIPEA, followed by **ii)** the addition of **23** at r.t. in dry DMF under Ar atmosphere, 18h, yielding <35% with tetramethyl urea impurities.

For the subsequent alkylation of the  $\gamma$ -phosphate of ATP, it was necessary to transfer ATP disodium salt into its tetrabutylammonium salt, in order to guarantee strictly anhydrous conditions.<sup>[98]</sup> The S<sub>N</sub>2 reaction afforded the  $\gamma$ -DBCO-PEG4-ATP (Scheme 2).



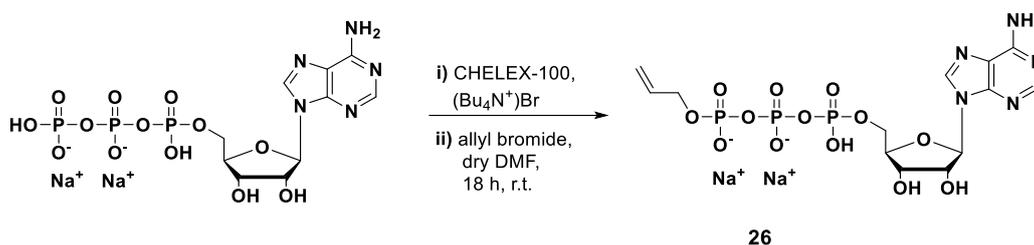
**Scheme 2.** Alkylation of the  $\gamma$ - phosphate of ATP with compound **24** via S<sub>N</sub>2 reaction, affording **25**.

In parallel to the synthetic approach of **25**, the  $\gamma$ -phosphate ATP analogues of the published compound library of Schönegger *et al.*, were screened towards their potential application as T4 PNK substrate.<sup>[87]</sup> These experiments revealed that rather short linkers between the clickable functionality and the  $\gamma$ -phosphate make the ATP analogues suitable as T4 PNK substrates, whereas ATP analogues with longer linkers did not afford successful labelling reactions with T4 PNK. This hypothesis was strengthened by molecular docking analysis.<sup>[87]</sup> Since compound **2**,  $\gamma$ -propargyl-ATP was identified as the best substrate for T4 PNK catalyzed 5'-end labelling reactions, we felt encouraged that  $\gamma$ -allyl-ATP, compound **26**, might be an equal substrate for T4 PNK, due to its identical linker length.<sup>[87]</sup> The idea was to label the 5'-OH of a ssDNA, RNA or oligonucleotides with compound **26** catalyzed *via* T4 PNK and subsequently perform a tetrazine ligation through an inverse electron-demand Diels–Alder reaction (IEDDA) with the allyl group of the potentially transferred  $\gamma$ -phosphate of compound **26**, which would enable another type of click reaction apart from CuAAC and SPAAC (Scheme 3).<sup>[87]</sup> The approach of using alkene-tetrazine ligation for labelling applications, where the nucleobases carry a vinyl functionality, has been published earlier and thus encouraged our approach.<sup>[33, 103-104]</sup>



**Scheme 3.** Suggested application of compound **26** for T4 PNK catalyzed labelling reaction and subsequent tetrazine ligation, with a tetrazine bearing molecule of interest in an inverse electron demand Diels Alder reaction (IEDDA).

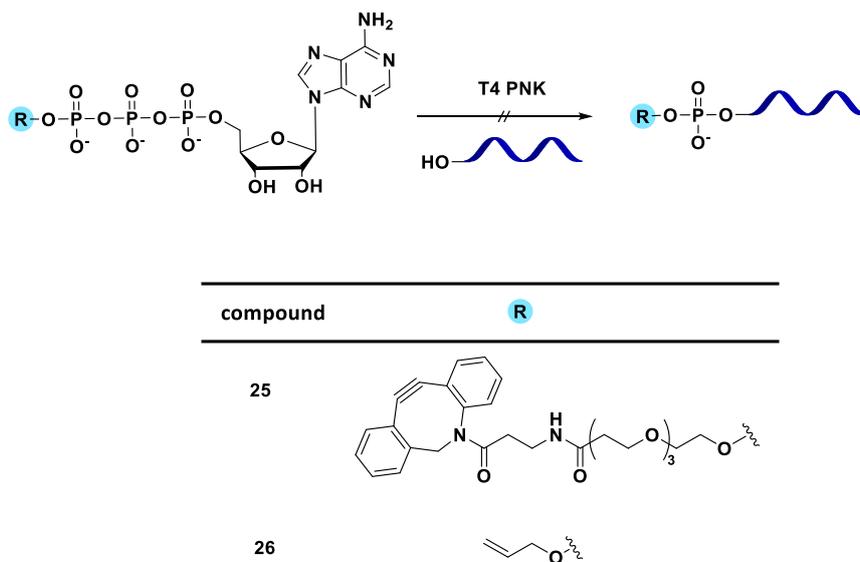
The synthesis of **26** was performed similar to **25**, with allyl-bromide as alkylating agent (Scheme 4).



**Scheme 4.** Synthesis of compound **26**.

Although the expectations for the T4 PNK catalyzed labelling reaction with compound **25** were low after previous experiments did not indicate successful labelling reactions with  $\gamma$ -ATP analogues bearing a long linker between phosphate and clickable functionality<sup>[87]</sup>, the T4 PNK catalyzed labelling reaction with a 43 mer oligonucleotide (ODN1) was nevertheless performed with  $\gamma$ -DBCO-PEG4-ATP (**25**). However, as expected, no labelling was observed, which appears reasonable, taking the long and bulky DBCO-PEG4 linker and the relatively narrow binding pocket of the T4 PNK into account.<sup>[105]</sup> The same T4 PNK catalyzed reaction with ODN1 was performed in the presence of compound **26**. To our surprise, despite the far less bulky end-group, no labelling of the 5'-OH of ODN1 was observed (Scheme 5). One reason therefore could be, that after the synthesis and 2-3 times RP-HPLC purification, inorganic salt impurities, such as bromide salts (e.g. NaBr), which are not visible *via* NMR, are still present. Those salt impurities might inhibit the subsequent T4 PNK catalyzed labelling reaction. However, as we isolated the sodium salts of the synthesized compounds after RP-HPLC purification by precipitation, we assumed to get rid of any other salt impurities. Previously we already faced similar issues where we observed higher labelling yields with compound **5**,  $\gamma$ -butynyl-ATP, which has been synthesized with the corresponding tosylate as alkylating agent, compared to the labelling with **5**, which has been synthesized with the corresponding bromide

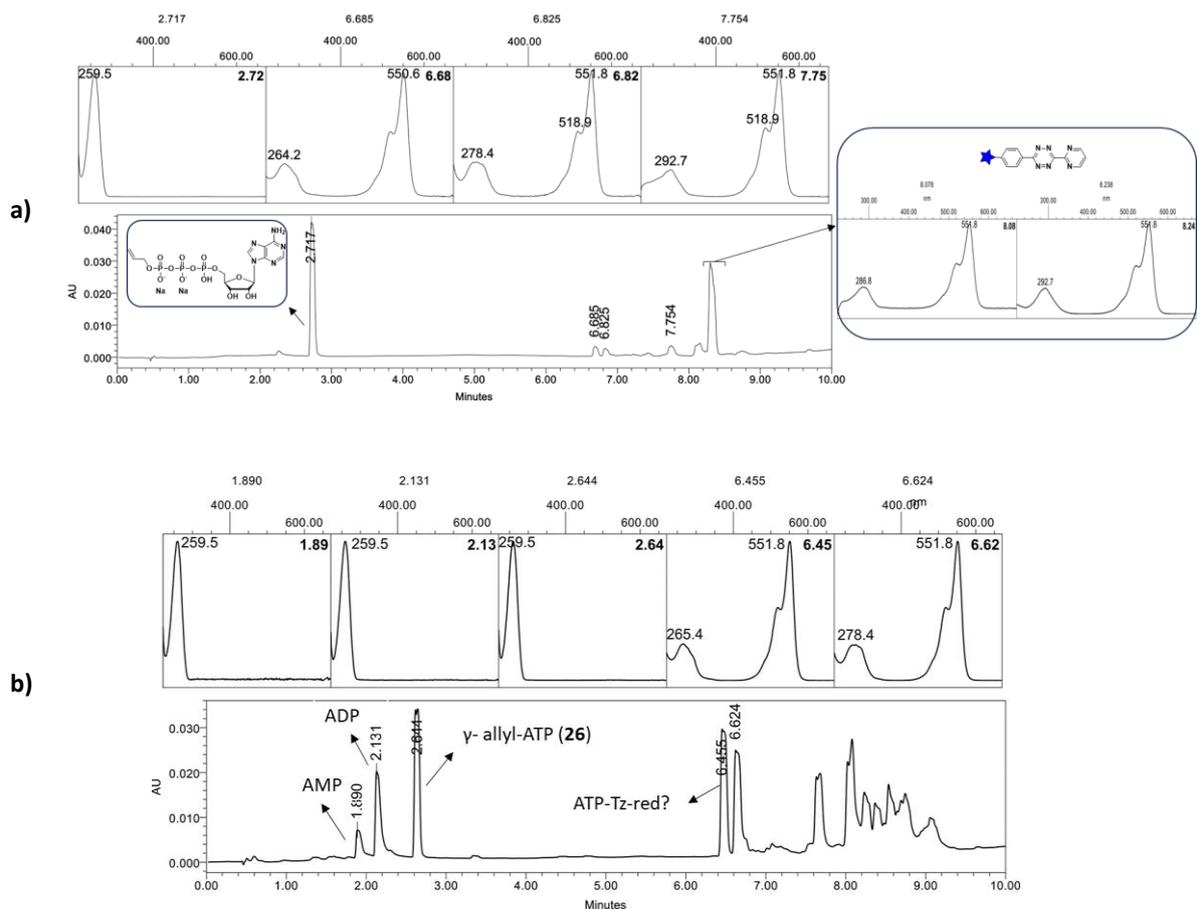
as alkylating agent. Thus we assume that although we performed several RP-HPLC purifications, followed by precipitation of compound **26**, remaining salt impurities, which were not visible by MS, HPLC and NMR, inhibit the T4 PNK reaction. To prove this hypothesis, it would be necessary to perform the synthesis of **26** with allyl tosylate, instead of allyl-bromide as alkylating agent and subsequently repeat the T4 PNK catalyzed labelling reaction with ODN1.



**Scheme 5.** Unsuccessful T4 PNK catalyzed labelling reaction on ODN1 with compound **25** and **26** respectively.

Since the original purpose of compound **26** was to perform a tetrazine ligation upon successful 5'-end labelling with the allyl group transfer catalyzed by T4 PNK, we anyway wanted to investigate whether compound **26** itself is suitable to undergo a tetrazine ligation, despite the labelling did not work. Therefore, we set up a reaction between compound **26** and a tetrazine bearing dye (Pyrimidyl-Tetrazine-Cy3, from Jena Bioscience), and tracked a potential conversion of the formed product by the UV trace in the RP-HPLC reaction control, indicating two absorbance maxima, one at 260 nm for the ATP analogue **26** and another one for the tetrazine dye at 550 nm.<sup>[106]</sup> Since tetrazine ligations with vinyl modified nucleobases have shown low second-order reaction rate constant with commercially available tetrazines, we did not expect our reaction to be fast and also assumed, based on previous tetrazine ligations with unstrained alkenes, the formation of more than one product, namely the oxidized, pyridazine (ATP-Tz-oxi) product and two tautomeric isomers of the reduced, dihydropyridazine (ATP-Tz-red) product (Scheme 6) <sup>[33, 103-104]</sup>

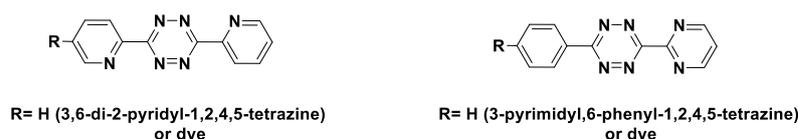




**Figure 19.** Analytical RP-HPL chromatograms **a)** Reaction control after 41 h. **b)** Reaction control after 75 h.

Unfortunately, due to time restrictions finishing the experiments for the publication, it was not possible to repeat the reaction in bigger scale for isolating sufficient material and performing subsequent  $^1\text{H}$  and  $^{31}\text{P}$  NMR analysis. Thus, it is still not clear whether the reaction worked or not. It would be necessary to repeat the reaction in at least 2 mg scale of compound **26** and a tetrazine bearing molecule, whereas it would be better to choose a simple tetrazine functionalized compound such as the commercially available 3,6-di-2-pyridyl-1,2,4,5-tetrazine (Figure 20), since it is cheaper and available in bigger amounts. This dipyridyl-s-tetrazine might also be beneficial for the reaction with the unstrained alkene, since previous work has shown successful IEDDA reactions with these tetrazine derivatives in combination with unstrained vinylboronic acids and unstrained S-allyl cysteine amino acid.<sup>[32, 107]</sup>

Originally the tetrazine-Cy3 dye was chosen in order to perform the click reaction on oligonucleotides, after the labelling *via* T4 PNK, with a similar dye as used for the CuAAC and SPAAC (Cy3-azide and DBCO-Sulfo-Cy3) in our previous publication.<sup>[87]</sup>



**Figure 20.** Structure of 3,6-di-2-pyridyl-1,2,4,5-tetrazine (derivate) (left). Structure of 3-pyrimidyl,6-phenyl-1,2,4,5-tetrazine (derivate) which has been applied for the performed IEDDA with compound **26** (right).

In case the tetrazine ligation does not work, which could be reasonable due to the rather low electron donating character of the allyl-phosphorester group, compound **26** could potentially undergo a thiol-ene reaction, which would enable yet another click reaction.<sup>[108-109]</sup>

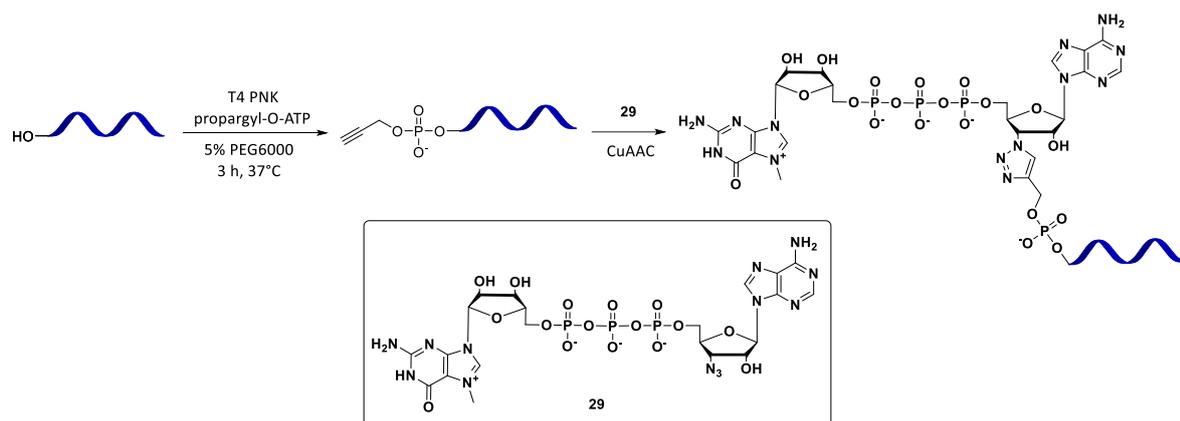
Although the synthesized  $\gamma$ -phosphate modified ATP analogues **25** and **26**, as well as some compounds from the work of Schönegger and co-workers<sup>[87]</sup>, were not suitable as substrates for T4 PNK, these compounds might be accepted by other enzymes. Serdjukow *et al.* for example showed, that  $\gamma$ -alkyne modified deoxynucleotide triphosphates, which have been linked to a fluorophore *via* click chemistry, can be incorporated into DNA by DNA polymerase.<sup>[73]</sup> This approach could be applied in single-molecule real-time (SMRT) sequencing.<sup>[110]</sup> Hence, it would be interesting to investigate whether a similar approach is possible with our  $\gamma$ -modified ATP analogues, in particular studying the incorporation of our fluorophore-clicked  $\gamma$ -modified ATPs in RNA by RNA polymerases, such as T7 RNA polymerase. These compounds could further be used to investigate protein kinase activities and other enzymatic processing.<sup>[111-113]</sup>

### Click-functionalized 5'-cap analogue

Encouraged by the successful 5'-end labelling of oligoribonucleotides<sup>[87]</sup>, we were motivated to expand the application tool box by designing a clickable 5'- cap analogue. We envision to chemically ligate a click-functionalized cap-derivate to the 5'- end of propargyl-labelled ribo nucleic acid, which has previously been installed by T4 PNK.

Motivated by the recent developments of 5'- cap analogues, we wanted to combine the approach of using synthetic anti-reverse cap analogues with the idea of modifying caps with a clickable functionality. Therefore, our m<sub>7</sub>GpppA cap (**29**) should bear an azide functionality directly at the 3'- position of A and after the successful 5'- end labelling of the 5' OH group of

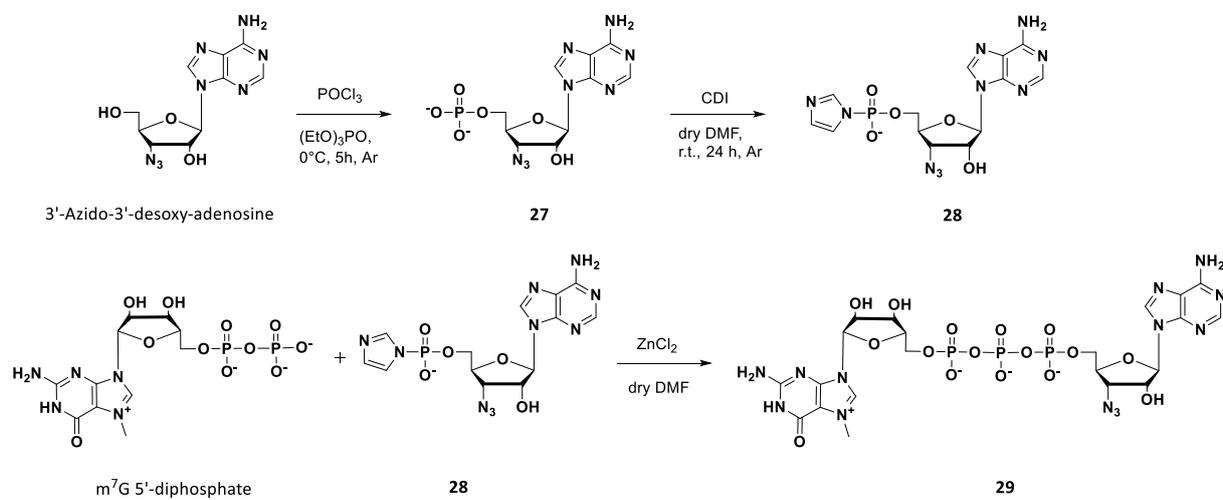
a synthetic oligoribonucleotide or a dephosphorylated IVT ribonucleic acid with  $\gamma$ -propargyl-ATP **29**, it should be attached *via* CuAAC to the transferred propargyl-phosphate of **29**, extending the native phosphodiester linkage between two nucleotides only by one 1,4-disubstituted 1,2,3-triazole linkage (Figure 21). This approach could be beneficial for post-synthetic capping of IVT RNA, or for capping of synthetic oligoribonucleic acids and subsequent investigation of diverse biological mechanisms and applications.



**Figure 21.** Proposed application of click-functionalized cap **29**. 1<sup>st</sup> step: 5'- end labelling of 5-OH of ribonucleic acid with  $\gamma$ - propargyl-ATP; 2<sup>nd</sup> step: CuAAC with the transferred propargyl-phosphate of **29**.

The proposed synthetic pathway for cap **29** (Scheme 7) is based on previous publications, where the first step, the phosphorylation of the commercially available 3<sup>1</sup>-azido-3<sup>1</sup>-desoxy-adenosine with POCl<sub>3</sub> was performed using an adjusted protocol from Stepinski and Zhang *et al.*<sup>[92, 114]</sup> The intermediate **27** was isolated after RP-HPLC as its tris-triethylammonium salt. The subsequent activation of the monophosphate with carbonyl diimidazole (CDI), resulted in a conversion yield of 70%. The typically higher yielding approach for the formation of phosphorimidazolide using an imidazole/triphenylphosphine/2,2'-dithiopyridine reagent system had to be avoided in order to prevent the 3<sup>1</sup>-N<sub>3</sub> undergoing a Staudinger reaction in the presence of triphenylphosphine.<sup>[115]</sup> After semi-preparative RP-HPLC purification, the collected product fractions indicated the desired mass of 421 in low resolution ESI-MS. Upon lyophilization, the characterization by high resolution ESI-MS and NMR (<sup>1</sup>H and <sup>31</sup>P) only revealed the starting material **27** (*m/z* 371) and the dimer GppG (*m/z* 725). We assume that **28** got hydrolyzed during lyophilization or storage in aqueous solution, which further led to the dimer formation. Since it was unfortunately outside of the experimental and funding time, it was not possible to repeat the reaction. However, we assume that it would be necessary to

isolate the product by precipitation, as described by Zhang *et al.*, instead of purification by RP-HPLC.<sup>[114]</sup>



**Scheme 7.** Proposed synthetic pathway for the preparation of the click-functionalized cap analogue **29**.

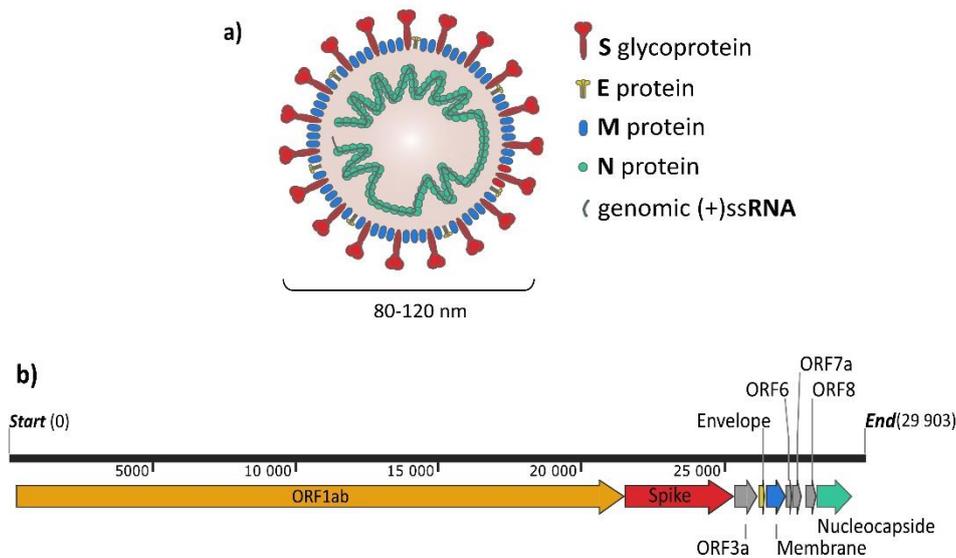
## 4.3 Click Modified Cu(II)-Dependent Nucleic Acid-Based Artificial Ribonuclease for Targeted Cleavage of SARS-CoV-2 Genomic RNA

### 4.3.1 Prologue

The coronavirus disease 19 (COVID-19) and its associated pandemic was caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which has first been identified in December 2019 in Wuhan (China), after clusters of patients with unknown caused pneumonia have been reported.<sup>[116-117]</sup> Due to its higher human-to-human transmission rate, compared to related beta coronaviruses (SARS-CoV, severe acute respiratory syndrome coronavirus and MERS-CoV, middle east respiratory syndrome coronavirus), SARS-CoV-2 accelerates a rapid spread across the world.<sup>[118]</sup> As of September 2023, there are more than 760 million confirmed positive cases and 6.9 million reported deaths worldwide.<sup>[119]</sup> Although as of June 2023, over 13 billion vaccines have been administered to decrease positive cases and deaths, there is still an imbalance of access to vaccines over the global population.<sup>[119-120]</sup> Moreover, the mutated variant strains of SARS-CoV-2 are still causing new local and global outbreaks by evading the immunity of vaccinations or previous infections.<sup>[118, 121-122]</sup> On the spike (S) protein alone, which the SARS-CoV-2 is utilizing for binding to the host cell surface ACE2 (angiotensin-converting enzyme 2) during the host cell entry and which is targeted by most vaccines, more than 80 variants have been identified.<sup>[123-124]</sup> Several “variants of concern”, including 1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), B.1.427/B.1.429 (Epsilon), B.1.617 (Kappa and Delta), and B.1.1.529 (Omicron) are harboring mutations in the S-protein, which facilitates to evade immunity and partially vaccine and monoclonal antibody therapeutics.<sup>[122, 125]</sup> Although the incidence of Covid-19 has been vastly decreased by vaccination in developed countries<sup>[126]</sup> and the World Health Organization (WHO) ended its public health emergency of international concern (PHEIC) declaration in May 2023 <sup>[127]</sup>, there is still an urgent need for alternative therapeutic approaches, especially towards targeting more conserved regions of the SARS-CoV-2 RNA in order to inhibit viral infection and progression.<sup>[128]</sup>

SARS-CoV-2 is an enveloped, positive-sense, single stranded RNA ((+)ssRNA) virus, bearing a genome of almost 30 000 nucleotides. It shares approximately 79% nucleotide sequence identity with SARS-CoV and about 50% with MERS-CoV.<sup>[129-130]</sup> It is composed of a 5'-untranslated region (UTR), the ORF1a/b (open reading frame), which encodes non-structural viral proteins, and a 3'-segment encoding for structural proteins, namely the spike (S),

nucleocapsid (N), envelope (E), and membrane (M) protein, and a 3'-UTR (Figure 22).<sup>[122, 130-131]</sup> While the S protein is responsible for entering the host cell by attaching to the ACE2 receptor, the N protein is supposed to have multiple functions, being mainly responsible for the identification and the wrapping of the viral RNA into a symmetrical structure.<sup>[132-135]</sup> The M protein, displaying the most abundant viral structural protein is essential for the virus assembly and budding, whereas the E protein is supposed to be multifunctional and is related to the pathogenesis process.<sup>[136-138]</sup>



**Figure 22.** a) Schematic drawing of SARS-CoV-2 virion of a diameter between 80-120 nm <sup>[139]</sup>, displaying the 4 structural proteins, S (spike), E (envelop), M (membrane), N (nucleocapsid). b) Map of the SARS-CoV-2 genomic RNA (NCBI RefSeq: NC\_045512.2 <sup>[117]</sup>)

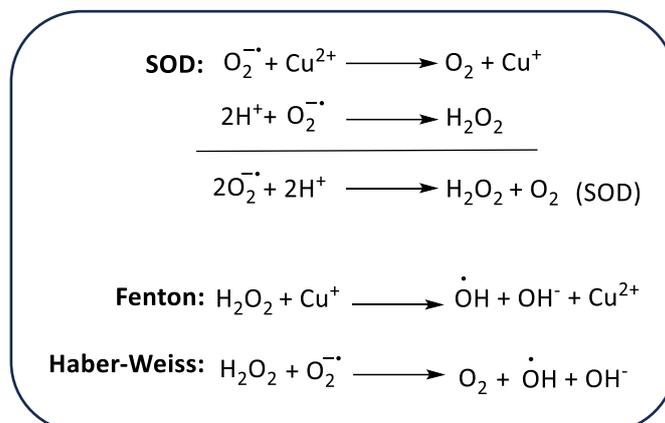
To stop the pandemic and further spreading of the virus, numerous SARS-CoV-2 vaccine candidates have been developed. As of December 2022, the WHO has granted 12 vaccines emergency use, whereof two of them, Spikevax from Moderna and Comirnaty from Pfizer/BioNTech are of RNA origin.<sup>[140]</sup> All vaccines are administered in order to confer immunity against a SARS-CoV-2 infection and thus preventing severe sickness and symptoms.<sup>[140]</sup> Nucleic acid-based vaccines, namely either DNA or RNA encoding for antigenic proteins, are using their hosts transcriptional and/or translational apparatus for producing disease specific antigens, namely proteins which are then processed, presented, and recognized by the immune system, thus activating T cell and humoral immune response.<sup>[141-142]</sup> While especially mRNA based vaccines showed great potential during the SARS-CoV-2 pandemic, the highly mutating nature of the spike protein, which has been the target of the

two first granted mRNA based vaccines, facilitates immune evasion and also monoclonal antibody resistance.<sup>[122, 141, 143]</sup> To overcome these challenges, several other therapeutic nucleic acids (TNA) are under development, including antisense oligonucleotides (ASOs) targeting the viral RNA.<sup>[122, 141, 144-145]</sup> Those short single-stranded nucleic acids target specific complementary sequences of the host RNA and depending on their design usually either induce target degradation upon RNase H recruitment (gapmer) or block translation by high-affinity binding to specific regions of complementary RNA (mixmer).<sup>[122, 146]</sup>

Hence, motivated by successful ASO therapeutics, targeting a variety of infectious, metabolic, and neurological human diseases <sup>[141]</sup>, and promising ASO candidates targeting the RNA of SARS-CoV-2 <sup>[122, 144-145]</sup>, the baseclick GmbH together with the LMU Munich and Kellett group from the Dublin City University), were eager to develop a novel antisense oligonucleotide related therapeutic to inhibit viral SARS-CoV-2 replication.

The idea of our ASO-related therapeutic is based on early developments in an artificial gene editing technology, namely artificial (ribo)nucleases. These structures consist of an oligonucleotide which is covalently linked to a “molecular scissor” moiety (organic or inorganic catalysts for RNA cleavage) and together these nucleic acid-metal complex hybrids enable catalytic cleavage of the RNA target, based on sequence specific target recognition.<sup>[147-150]</sup> These “molecular scissors” are often specifically designed ligands, which upon complexation of metal ions (often divalent metal ions, such as divalent transition metal ions Zn(II) and Cu(II), which play an important role in the catalytic mechanism), induce artificial nuclease activities.<sup>[148-149]</sup> The Kellett group has applied this approach towards chemically modified triplex forming oligonucleotides (TFO), bearing a Cu(II) binding ligand, which, upon coordination of copper ions, leads to sequence specific cleavage.<sup>[147, 151]</sup> The mechanism of the strand cutting is based on the *in-situ* formation of a Cu(II)-ligand-TFO-target hybrid, which upon the addition of a reductant, such as ascorbic acid or superoxide radicals ( $\bullet\text{O}_2^-$ , Figure 23)<sup>[152]</sup>, creates the active Cu(I)-ligand-TFO-target-hybrid. Cu(I) further leads to the formation of reactive oxygen species (ROS), especially hydroxyl radicals ( $\bullet\text{OH}$ ) *via* Fenton/Haber–Weiss reaction (Figure 23), which is ostensibly the most powerful oxidizing radical in biological systems.<sup>[152-156]</sup> The necessary hydrogen peroxide ( $\text{H}_2\text{O}_2$ ) is generated *via* superoxide dismutase (SOD) type pathway, from superoxide radicals ( $\bullet\text{O}_2^-$ ). These ROS are then supposed to act as chemical

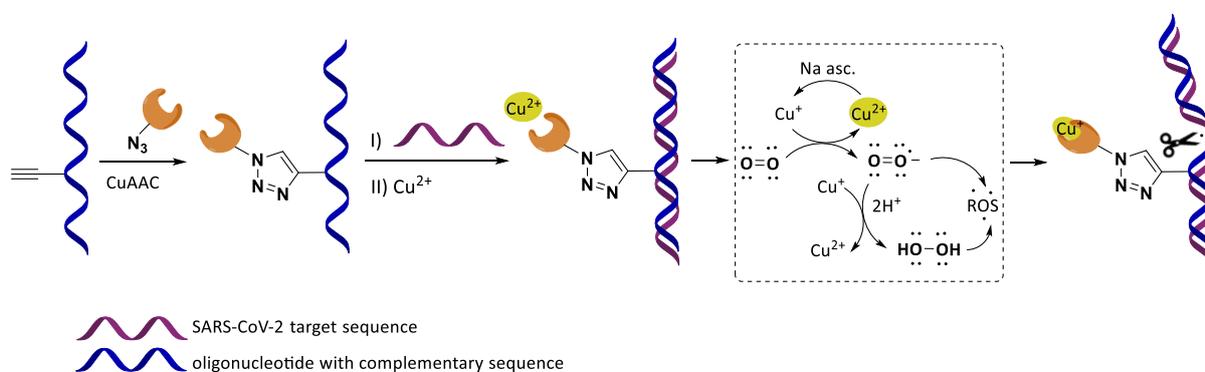
nuclease.<sup>[151, 157-158]</sup> The Kellett group further developed this artificial gene editing approach by attaching metal-binding ligands to TFOs using click chemistry.<sup>[151, 159-160]</sup>



**Figure 23.** Formation of radical oxygen species, by the participation of Cu ions <sup>[154-156]</sup> SOD= superoxide dismutase (SOD) type pathway.

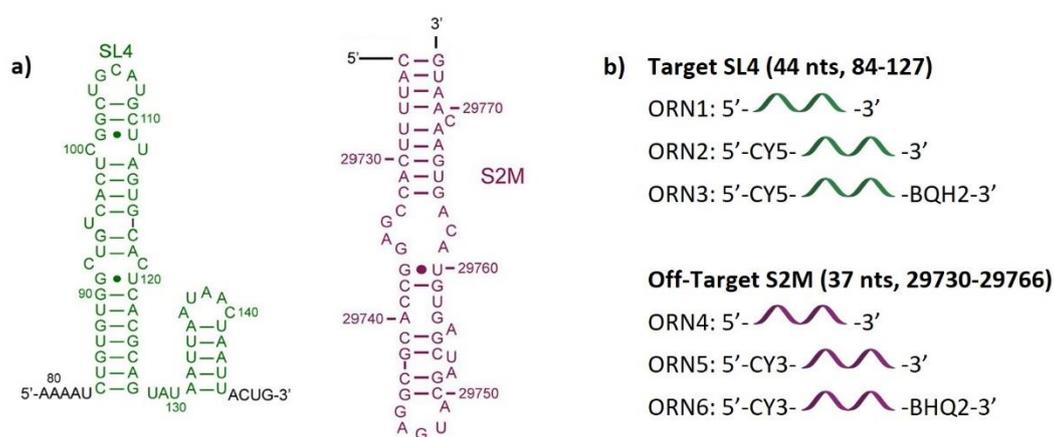
#### 4.3.2 Results and Discussion

In our novel approach, we now wanted to combine the method mentioned above with classic ASO therapeutic technology. Herein, a chemically modified oligonucleotide bearing a Cu-binding ligand, targeting a highly conserved stem loop sequence of the SARS-CoV-2 genomic RNA, should, upon binding, form a DNA-RNA hybrid. After complexation of copper and reduction of Cu(II) to Cu(I), the thereby formed ROS should preferably lead to site-specific cleavage of the SARS-CoV-2 RNA and thus inhibit viral replication (Figure 24). For initial experiments, Cu(II) is added upon hybrid formation, however for specificity reason, we later aim to pack the ASO system, including Copper, into lipid nanoparticles.



**Figure 24.** Schematic representation of our proposed mechanism of site-specific cleavage of SARS-CoV-2 genomic RNA.

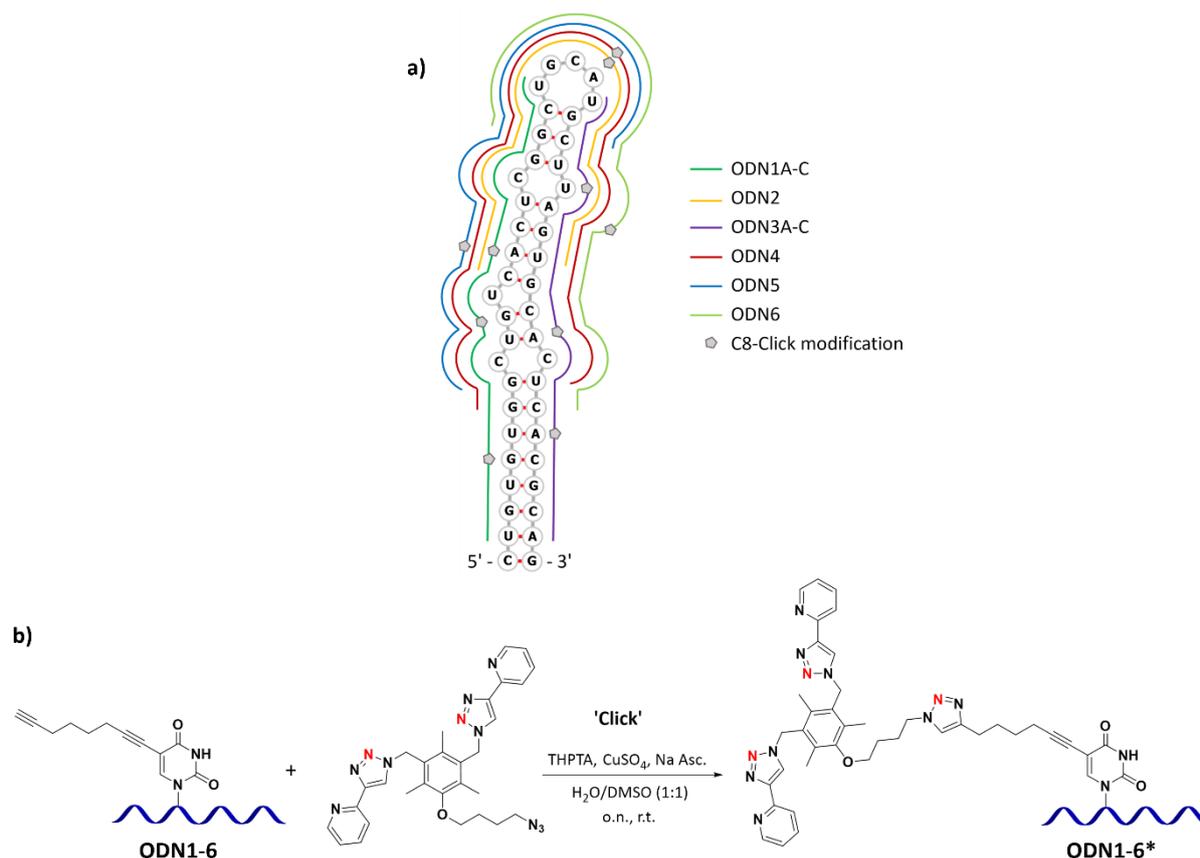
In order to overcome potential limitations of our approach caused by viral mutations, we aimed to target the stem-loop 4 (SL4) sequence of the SARS-CoV-2 genomic RNA. SL4 is one of 5 hairpin loop structures of the highly structured 5'-UTR and is supposed to be a relatively stable and long hairpin. It has further been hypothesized that SL4 participates in the synthesis of subgenomic RNA fragments.<sup>[161-163]</sup> For the off-target sequence, stem-loop 2 motif (S2M) at the 3'-UTR was selected. (Figure 25a).<sup>[164]</sup> In order to investigate binding specificity and cleavage activity towards the target and off-target using different techniques, a series of three oligoribonucleotides for each, target and off target have been designed. ORN1 and ORN4 represent the unmodified sequences of target and off-target sequences respectively. ORN2 and ORN5 are bearing a 5'-fluorescent dye, namely SulfoCy5 and Cy3 respectively, whereas ORN3 and ORN6 display, additionally to the 5'-dye modification, a 3'-BHQ2 (black hole quencher 2-phosphate) modification (Figure 25b). The modifications of ORN3 and ORN6 were designed in reference to Farzan *et al.*<sup>[165]</sup>



**Figure 25. a)** SL4 and S2M stem loop sequences of the SARS-CoV-2 genomic RNA; images taken from Ryder *et al.*<sup>[163]</sup> **b)** Target and off-target sequences mimicking the SL4 and S2M stem loop respectively.

For first screenings, 10 oligonucleotide sequences, ODN1-ODN6, targeting the SL4 were designed in reference to Lulla *et al.*, covering different positions of the hairpin structure with varying sequence length between 19 and 31 nts (Figure 26).<sup>[164]</sup> In order to attach a Cu-binding ligand *via* click chemistry, the oligonucleotides were synthesized by standard phosphoramidite synthesis, incorporating one C8-alkyne modified base in each strand. In order to investigate the best binding affinity and target specificity, different positions of the internal C8-alkyne modification on varying bases have been chosen (Figure 26a). These C8-alkyne modified oligonucleotides were ligated to the Cu-binding “DC-Py-N<sub>3</sub>” ligand, using Cu(I)

catalyzed azide-alkyne cycloaddition (CuAAC), which was provided by the Kellett group. The novelty of this ligand, invented by the Kellett group is, that once it is clicked to an oligonucleotide, the formed 1,4-disubstituted 1,2,3-triazole, in particular one nitrogen atom thereof (Figure 26b), takes actively part in the copper coordination.

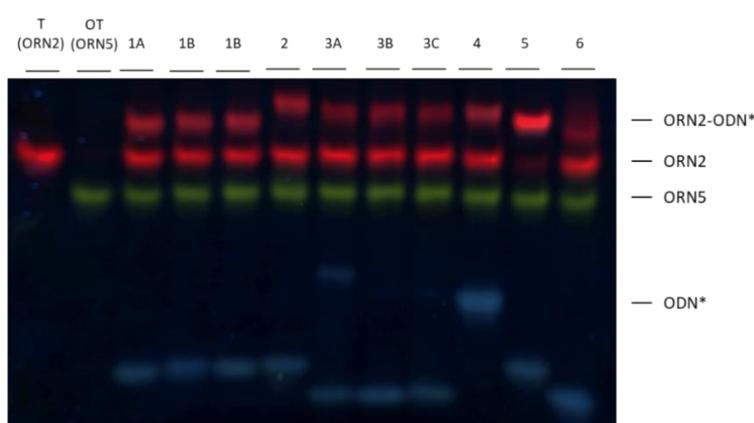


**Figure 26.** **a)** Schematic representation of target ORN1, mimicking SL4 (84-127 nt), with the annealing, complementary oligonucleotide sequences ODN1-6. For illustrative purposes, the depicted secondary structure was calculated using RNAfold web server<sup>[166]</sup> using the default parameters, and then visualized using forna<sup>[167]</sup>. **b)** Scheme of CuAAC click reaction of C8-alkyne modified oligonucleotides ODN1-6 with the DC-Py-N<sub>3</sub> Cu(I) binding ligand, yielding the Cu(I)-ligand binding oligonucleotides ODN1-6\*. For exemplary illustration, an internal C8-alkyne dT modification is displayed. The nitrogen atoms, in red are actively taking part in the Cu coordination.

The CuAAC of the C8-alkyne modified oligonucleotides ODN1-ODN6, were performed in two different scales, namely 25 nmol for test reactions and 250 nmol for scale up. Herein, CuSO<sub>4</sub> was used as Cu(II) source, which got reduced *in-situ* by sodium ascorbate to form the active Cu(I) species, being necessary for the catalytic cycle.<sup>[10]</sup> To maintain Cu(I) in its active form and thus accelerating the click reaction, THPTA (tris(3-hydroxypropyltriazolylmethyl)amine) was used as ligand during the click reaction and once no further conversion of the reaction was monitored *via* RP-HPLC, the reaction was quenched with EDTA in order to complex the Cu

ions.<sup>[151, 168]</sup> Prior to semi-preparative RP-HPLC purification, the crude reaction mixtures were purified by gel filtration (size exclusion chromatography). After RP-HPLC, purification the Cu-binding ligand clicked oligonucleotides were again desalted, by gel filtration (size exclusion chromatography), in order to remove residual buffer (TEAA) impurities from the RP-HPLC, which could interfere with subsequent experiments.

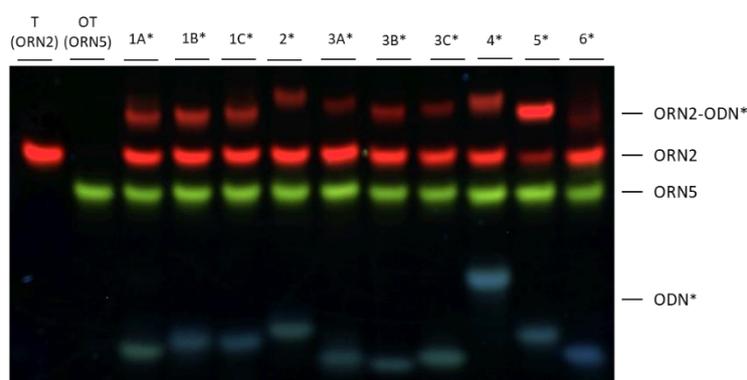
For the initial analysis of the binding affinity and target specificity of the oligonucleotides towards target and off-target sequences, all C8-alkyne modified oligonucleotide (ODN1-ODN6) were incubated individually with target and off-target oligoribonucleotides, bearing the 5'-dye modification (ORN2 and ORN5), in a ratio of 1:5 (ORN:ODN). The samples were incubated at physiological conditions (37°C, pH 7) for 2 h and subsequently analyzed *via* native polyacrylamide gel electrophoresis (nPAGE). The gels were imaged using different filters, according to the fluorescent dyes, subsequently stained with SYBR Gold and then imaged again using additionally the blue filter to display the SYBR Gold staining. The gel image (Figure 27) indicates that ODN5 appears to be the best binder for ORN2 since it shows the highest intensity of the ORN2-ODN hybrid and only a very weak signal for ODN2 alone. ODN6 on the other hand does not show high binding affinity towards ODN2, since the signal of ODN2 alone is relatively high compared to 5. ODN1 (A-C), ODN2, ODN3 (A-C) and ODN4 indicate also binding affinity towards ODN2, but with a lower intensity level compared to ODN5. No ODN showed binding affinity towards the off target.



**Figure 27.** nPAGE (20%) of C8-alkyne modified oligonucleotides (ODN1-6) incubated with ORN2 (T) and ORN5 (OT); image was taken after SYBR Gold stain; filter: red= Cy5, green=Cy3, blue= SYBR Gold.

Next, the Cu-binding ligand clicked oligonucleotides (ODN1-6\*) were incubated with target ORN2 (T) and off target ORN5 (OT) as described above to investigate if and how the Cu-

binding-ligand changes the pattern of the binding affinities towards target and off target. The nPAGE and imaging was performed as described above and the gel image is displayed in Figure 28. The gel image of the ODN1-6\* indicates a similar pattern as the C8-alkyne modified oligonucleotides. Again, ODN5\* has the best binding affinity towards the target (ORN2), while ODN6\* does not show any ORN2-ODN6\* hybrid at all. ODN1\* (A-C), ODN2\*, ODN3\* (A-C) and ODN4\* form ORN2-ODN\* hybrid, however in a significant lower intensity compared to ODN5\*. Especially for ODN3\* (A-C) the hybrid formation seems to be very weak, which can be assumed from the low signal of the ORN2-ODN\* hybrid. Since, both binding analyses, with C8-alkyne and Cu-binding ligand clicked oligonucleotides revealed that ODN5(\*) shows the highest affinity towards the target, while ODN6(\*) does not and ODN3 (A-C) especially for the clicked oligonucleotides gave a very poor signal for the hybrid, the following analysis regarding binding and cleavage were performed only with the sequences of ODN5\*, ODN4\*, ODN2\* and ODN1\* (A-C).

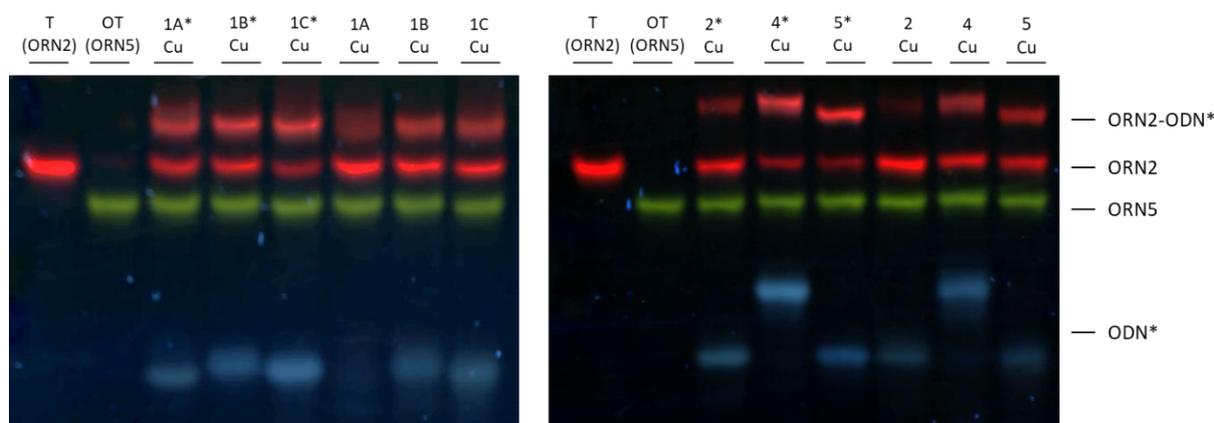


**Figure 28.** nPAGE (20%) of Cu-binding ligand clicked oligonucleotides (ODN1-6\*) incubated with ORN2 (T) and ORN5 (OT); image was taken after SYBR Gold stain; filter: red=Cy5, green=Cy3, blue= SYBR Gold.

Since the crucial step of our approach is based on the complexation of copper by the ligand clicked on the oligonucleotide sequences, the next binding analysis was performed with a copper source, namely  $\text{Cu}(\text{Cl}_4\text{O})_2$ . Therefore, the selected Cu-binding ligand-ODN (5 equiv. of target ODN) was mixed with an aqueous  $\text{Cu}(\text{Cl}_4\text{O})_2$  solution (10 equiv.) and incubated at r.t. for 10 min. Afterwards target and off target (ORN2 and ORN5; 1 equiv. each) were added and incubated for 18 h at 37°C. For comparison also the C8-alkyne modified ODN were incubated with the copper source and ORN2 and ORN5 in the same manner. As reference, target and off target samples alone were also incubated under the same conditions (buffer, temperature, time). Subsequent nPAGE and imaging was performed as described above. The gel images are

displayed in Figure 29. It reveals that ORN2-ODN\* hybrids, especially with ODN4\* and ODN5\* show high intensity levels, followed by ODN1\* (A-C). ODN2(\*) in contrast, showed only poor hybrid formation. Moreover, the direct comparison of ODN4/5 and ODN4\*/5\* with copper indicated, that upon copper complexation the ORN2-ODN\* hybrid even seems to be stabilized, which would be beneficial for specificity of subsequent cleavage.

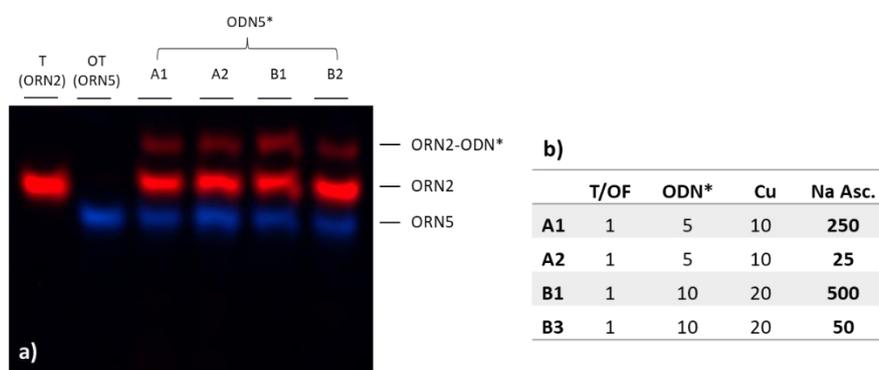
Overall, the binding assays reveal that the oligonucleotide sequences annealing at the position towards the 5'-end of the target stem loop (including the loop itself), show a higher binding affinity towards the target. The finding, that the better affinity appears on the 5'-upstream region of the stem loop has already been shown in a previous publication.<sup>[164]</sup>



**Figure 29.** nPAGE (20%) of ODN(\*)-ORN binding analysis with and without copper (Cu); image was taken after SYBR Gold stain; filter: red= Cy5, green=Cy3, blue= SYBR Gold.

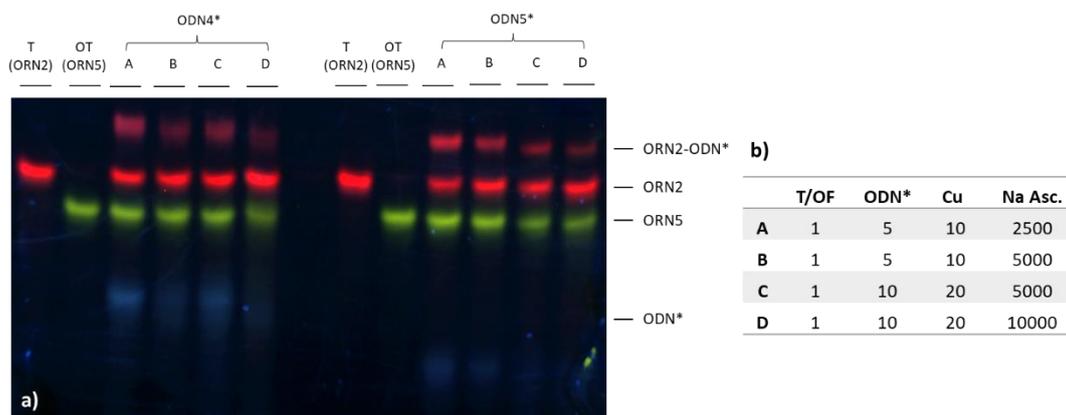
Since ODN4\* and 5\* indicated the best binding affinity towards target ORN2, the cleavage assays were performed only with these two sequences, whereas the first cleavage assay was performed solely with ODN5\* and the conditions were based on previous experiments using TFOs for artificial gene editing approaches.<sup>[159-160]</sup> For this first cleavage approach, ODN5\* was incubated with  $\text{Cu}(\text{Cl}_4\text{O})_2$  like in the previous experiment, before Na ascorbate was added and incubation continued at r.t. for 10 min. Finally, ORN2 and ORN5 were added and incubated at 37°C for 18 h. The following nPAGE and imaging was performed as previously described and the gel image before SYBR Gold staining, as well as the different conditions are displayed in Figure 30. Looking at the gel image however indicates that no degradation of any band can be detected, thus the amount of sodium ascorbate in the reaction mixture might have been too low to lead to successful Cu(II) reduction and subsequent Cu(I) induced cleavage. On the other hand, the ORN-ODN band intensity in the gel image is weaker than in the previous

experiments, thus maybe there was even some degradation, but we were not able to detect the degradation products, by accident forgot to perform SYBR Gold staining, as we assumed to expect a smear rather than only weaker band intensities. Nevertheless, for the next experiment the amount of sodium ascorbate needed to be increased.



**Figure 30. a)** nPAGE (20%) of ODN5\* incubated with T/OF,  $\text{Cu}(\text{Cl}_4\text{O})_2$  and sodium ascorbate (Na asc.) according to the conditions, given in table **b)** (1= 1 equiv.= 0.5 pmol; details see section 6.4.4) image was taken after before SYBR Gold stain; filter: red= Cy5, blue= Cy3.

For the second cleavage experiment, the samples for ODN5\* and ODN4\* were prepared in the same manner as in the previous experiment, with modified molar ratios, especially with an increased ratio of copper to reductant (Figure 31b). The gel image (Figure 31a) now indicates clear degradation, which is attributed to the appearing smear of the ORN2-ORN4\* hybrids. The smear, however, does not only appear for the ORN2-ODN4\* hybrid, but also for the oligos themselves. Similar pattern applies for ODN5\*, yet in a lower extent. Especially the conditions B, C and D, with either 1:500 (B and D) or 1:250 (C) seem to lead to cleavage. The conditions seem to be better compared to the previous experiment, since this experiment showed evidence for cleavage. However, the damage does not indicate target specificity, which is mandatory for this approach. Also, it would have been necessary to perform a negative control (without the addition of copper and sodium ascorbate), to enable a direct comparison.

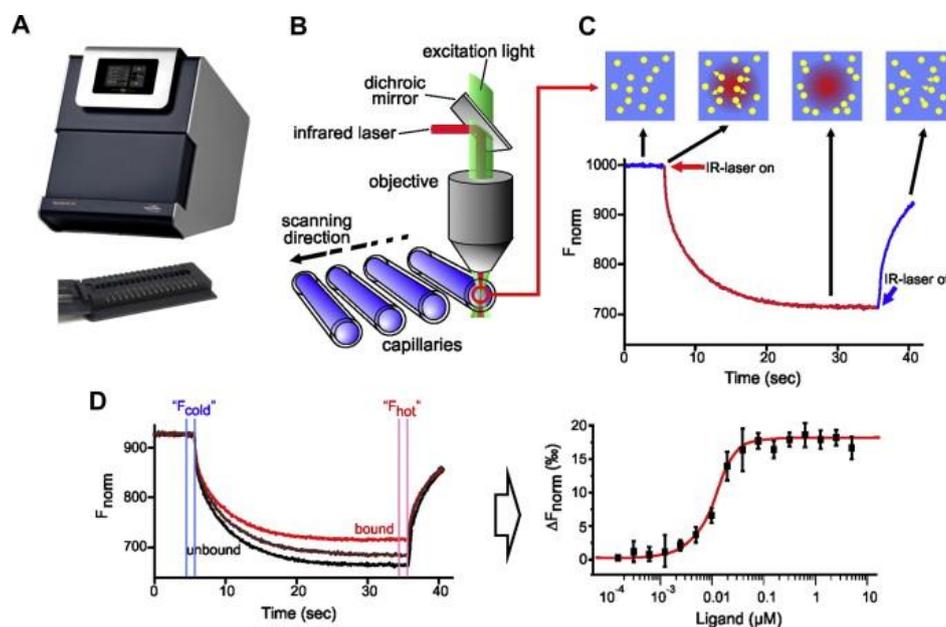


**Figure 31. a)** nPAGE (20%) of ODN4\* and ODN5\* incubated with T/OF,  $\text{Cu}(\text{Cl}_4\text{O})_2$  and Na ascorbate according to the conditions A-D, given in table **b)**; red= Cy5, green=Cy3, blue= SYBR Gold.

In order to quantify the interaction strength of the target oligoribonucleotide and best performing oligonucleotides, microscale thermophoresis (MST) experiments have been carried out.

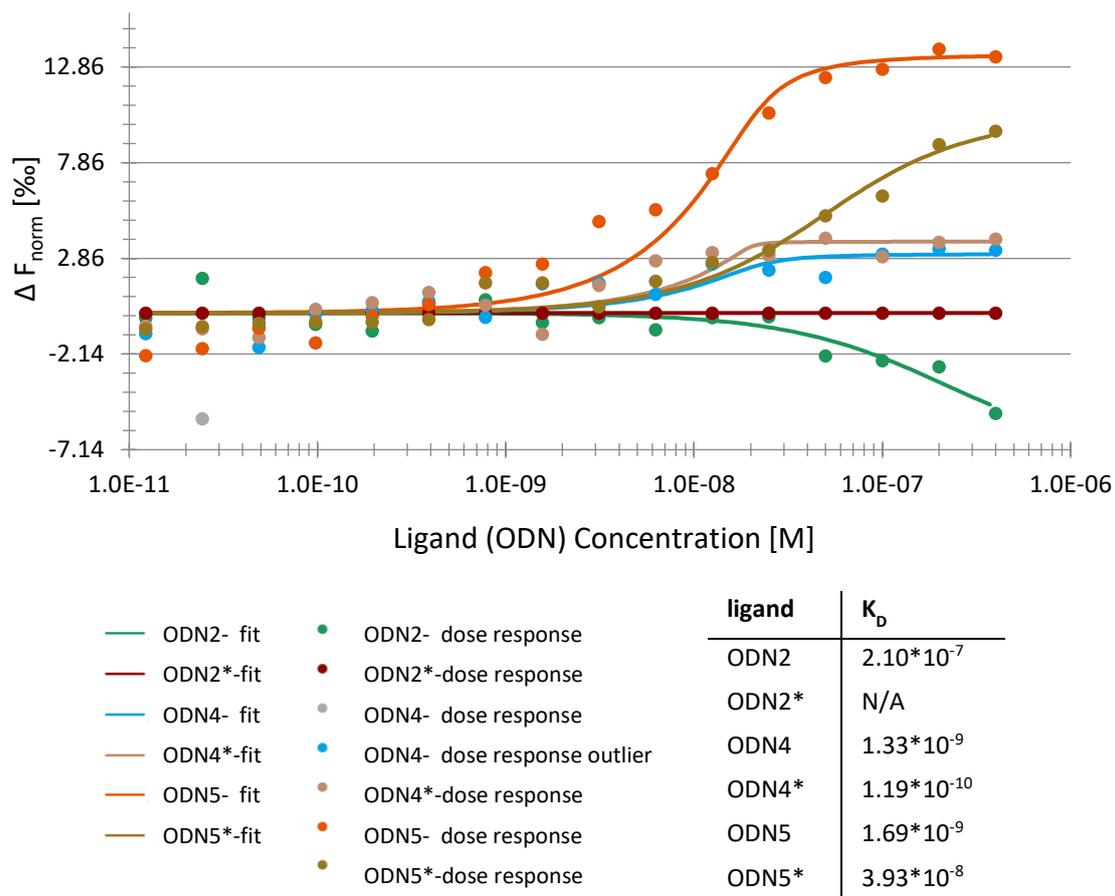
Excursion: Thermophoresis describes the directed movement of molecules in a temperature gradient and since this movement strongly depends on molecular properties (e.g. size, charge, conformation, hydration shell), this method is highly sensitive to changes in molecular properties. Once a desired molecule binds to the target molecule, the molecular properties will change, resulting in a change in the speed by which the target molecules escape from the temperature gradient. This temperature gradient ( $\Delta T = 2 - 6^\circ\text{C}$ ) is induced by an infrared laser ( $\lambda_{\text{emission}} = 1480 \text{ nm}$ ). For detecting and quantifying the directed movements of the molecules through this temperature gradient, which is based on mass diffusion, covalently attached or intrinsic fluorophores are used. The change of the thermophoretic movement, based on the Temperature-Related Intensity Change (TRIC) measurement, is then expressed as the change in the normalized fluorescence ( $\Delta F_{\text{norm}} = F_{\text{hot}}/F_{\text{cold}}$ ) and further plotted against the ligand concentration in logarithmic scale. The resulting binding curve, which is ideally of sigmoidal shape, is subsequently fitted in order to derive the binding constants.<sup>[169-171]</sup> The inflection point, which corresponds to the concentration of ligand at which half of the target molecules are bound to ligand, equal the dissociation constant  $K_D$ .<sup>[171-172]</sup> For a reversible reaction  $L+M \leftrightarrow LM$ , the dissociation constant is defined as  $K_D = \frac{[L] \times [M]}{[LM]}$ , thus the lower the  $K_D$  value is, the tighter the interaction and the higher the affinity between molecule and ligand are.<sup>[171]</sup> Since the  $K_D$  is directly depending on the Gibbs free energy change,  $\Delta G^\circ = RT \cdot \ln(K_D)$ , upon measuring the  $K_D$  values over a temperature range, thermodynamic parameters for biomolecular

interactions, can be derived ( $\Delta G^\circ$ ,  $\Delta S^\circ$ ,  $\Delta H^\circ$ ).<sup>[169]</sup> The typical MST experimental setup and data generation is displayed in Figure 32.



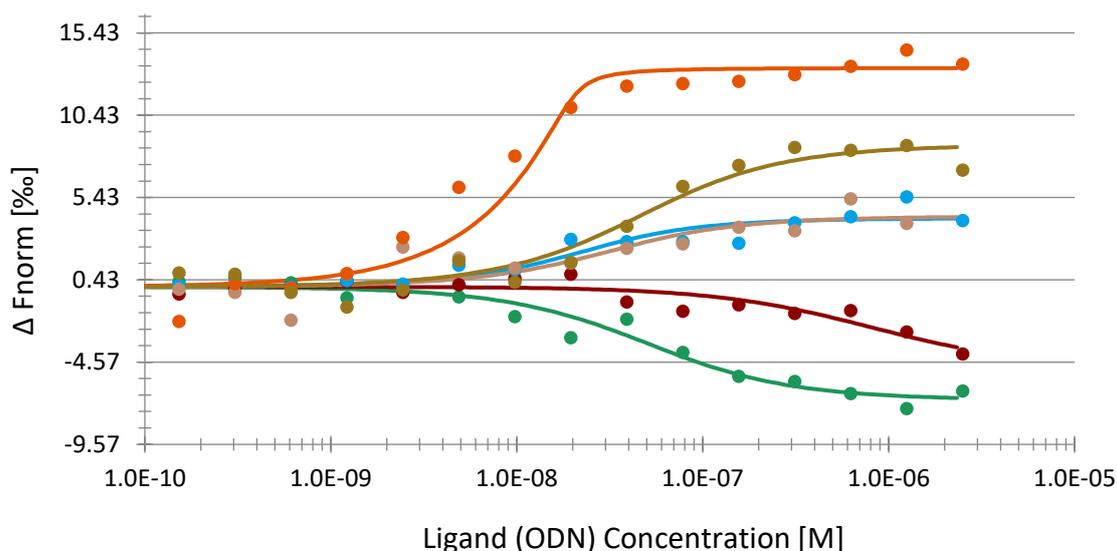
**Figure 32.** Figure taken from Jerabek-Willemsen *et al.*<sup>[169]</sup> without modification and with preprint permission from Elsevier. **A.** Monolith NT.115 from NanoTemper Technologies GmbH and the capillary tray (image below). **B.** Schematic representation of the optical setup for MST. The objective excited and detects the fluorescence within the capillaries (total volume of  $\sim 4 \mu\text{L}$ ), whereas a focused IR- laser heats a defined sample volume. **C.** Characteristic signal of a MST experiment. In the beginning, when the molecules are distributed homogeneously, a constant “initial fluorescence” is detected. Upon activation of the IR laser, the “T-Jump” (T=temperature) can be observed, indicating a rapid change in fluorophore properties, caused by the fast temperature change. Afterwards, the thermophoretic movement of the fluorescently labeled molecules out of the heated sample volume can be detected, whereas the fluorescence change is usually measured for 30 s. An inverse T-Jump occurs, upon the IR laser gets deactivated, which is followed by the “backdiffusion” of molecules, a process driven solely by mass diffusion. **D.** Characteristic binding experiment. Changes of the thermophoretic movement of the fluorescent molecule (“unbound” trace) upon binding to a non-fluorescent ligand (“bound”), result in different traces.  $\Delta F_{\text{norm}}$  (the change in the normalized fluorescence), which is defined as  $F_{\text{hot}}/F_{\text{cold}}$  (F-values are corresponding to average fluorescence values between defined areas marked by the red and blue cursors, respectively), expresses the change in thermophoretic movement (thermophoresis). Gradual change in thermophoresis, which is achieved by titration of the non-fluorescent ligand, is plotted as plotted as  $\Delta F_{\text{norm}}$ , yielding a binding curve, that can be fitted to derive binding constants.<sup>[169]</sup>

For our initial MST experiment, ORN2, bearing a 5'-SulfoCy5 fluorescent dye, was incubated with ODN2, 4, 5 and their corresponding Cu-ligand clicked sequences (ODN\*) in individual experiments at physiological conditions (4h, 37°C) in a ratio of 1:20 (ORN:ODN) and subsequently MST was performed. The fitted binding curves and the calculated  $K_D$  values are displayed in Figure 33.



**Figure 33.** Binding curve displaying the ODN(\*) (ligand) concentration ( $c_{max}=400$  nM) versus the  $\Delta F_{norm}$  and  $K_D$  values of the individual ligands.

Looking at dose-response curve in Figure 33, it can be observed, that the ligand concentration range within this experiment was not appropriate for all ligands. While for ODN5, ODN4 and ODN4\* the curve is of almost sigmoidal shape, for the other ODNs the plateau indicating binding of the target, is not reached. Thus, for ODN2\* no  $K_D$  values could be calculated and some of the calculated  $K_D$  values might be error prone since the bound stages, which are indicated by reaching a plateau, are not yet reached with this concentration range. Further, when looking at the almost sigmoidal shape for ODN5, it seems that actually two binding events are happening, thus the  $K_D$  fitting might not be appropriate, thus leading to inaccurate data. To investigate whether higher ligand concentration gives sufficient data, a second experiment was performed, where the ligand concentration was increased from a maximal concentration of 400 nM to 2.5  $\mu$ M, while keeping the target concentration 20 nM, resulting in a ratio of 1:250. The fitted binding curves and the calculated  $K_D$  values are displayed in Figure 34.



—	ODN2- fit	●	ODN2- dose response	<b>ligand</b>	<b><math>K_D</math></b>
—	ODN2*-fit	●	ODN2*-dose response	ODN2	$4.03 * 10^{-8}$
—	ODN4- fit	●	ODN4- dose response	ODN2*	$8.09 * 10^{-7}$
—	ODN4*-fit	●	ODN4- dose response outlier	ODN4	$1.18 * 10^{-8}$
—	ODN5- fit	●	ODN4*-dose response	ODN4*	$2.09 * 10^{-8}$
—	ODN5*-fit	●	ODN5- dose response	ODN5	$4.27 * 10^{-10}$
		●	ODN5*-dose response	ODN5*	$3.60 * 10^{-8}$

**Figure 34.** Binding curve displaying the ODN(\*) (ligand) concentration ( $c_{max}= 2.5 \mu\text{M}$ ) versus the  $\Delta F_{norm}$  and  $K_D$  values of the individual ligands.

The higher ligand concentrations lead to dose response curves with an almost sigmoidal shape for all ODNs and nearly reaching the plateau for all ODNs. Thus, all  $K_D$  values could be calculated using the  $K_D$ -fit model of MO Analysis software. Since the direction of the binding curve depends on the specific properties of the molecules, the direction does not influence the  $K_D$  analysis.<sup>[173]</sup> In this MST experiment, ODN5 revealed the lowest  $K_D$  value of 0.427 nM for the C8-alkyne modified oligonucleotides, while for the ligand clicked oligonucleotides, ODN4\* with a  $K_D$  of 20.9 nM, gave the lowest values, followed by ODN5\* with a  $K_D$  of 36.0 nM. For all individual sequences the C8-alkyne modified oligonucleotides show a lower  $K_D$  values compared to the Cu-ligand clicked oligonucleotides. The reason for this trend is very likely based on the sterically hinderance of the Cu-ligand, which would interfere with the base pairing of the ODN-ORN hybrid and thus weaken their binding. Considering the sequence length of the best performing oligonucleotides ODN4\* and ODN5\* with 31 and 20 nt respectively, the  $K_D$  values of ODN5\* are relatively low, if taking into account that ODN4\*

displays 9 nucleotides more which can contribute to the hybridization towards the target sequence. In general, considering the experimental setup and the generated data, it has to be questioned whether MST is an appropriate method for analyzing the binding affinities of target-ligand interactions. On one hand, the experiments are not performed at physiological conditions, whereas the actual binding event should take place at 37°C. Furthermore, the size of the ligand oligos might have an impact on the speed of moving out of the laser. Since once the laser is activated, the duplex structures melt and the oligos are moving out of the laser- spot depending on their size. On the cold side they could hybridize again, which process is influenced by the oligos that have entered the cold zone by that time, thus it might be inappropriate to neglect the length of the ligand oligos for MST. Nevertheless, the MST results coincide with the binding analysis *via* nPAGE, revealing that ODN4\* and ODN5\* as the most promising oligonucleotide candidates for the desired approach, due to their high affinity towards the target ORN2. For prospective analysis towards binding affinity, isothermal methods might be more suitable, where for example surface plasmon resonance (SPR), or isothermal calorimetry (ITC) could be used. <sup>[174]</sup>

In conclusion, some preliminary, but nevertheless encouraging results have been obtained towards the novel approach for targeted, oligonucleotide based, cleavage of the SARS-CoV-2 genomic RNA.

We envision that after a certain degree of optimization, site specific cleavage of the target sequence could be achieved. Apart from investigating different ratios of target, ODN\*, Cu-source and reductant, the linker length between the nucleobases and the Cu(II)-binding ligand could be modified towards shorter linkers, since the C8-linker might cause unfavored flexibility leading to unspecific cleavage. This could be achieved by replacing the C8-alkyne modifications with ethynyl modified bases, by incorporating TIPS-5-Ethynyl-dU-CE phosphoramidites during solid phase synthesis. However, the CuAAC reactivity of the ethynyl group in DNA is lower compared to the octadiynyl group, and the short linkage between the Cu-binding ligand and the oligonucleotide could interfere with the DNA-RNA base pairing, which might weaken the ODN\*-target affinity. <sup>[39], [175]</sup> Moreover, the Cu(II)-binding ligand itself could be replaced by other e.g Zn(II)-binding ligand. The use of Zn(II) artificial nucleases, namely Zn(II) binding PNA (peptide nucleic acid) conjugates (PNAzymes) have recently shown cleavage of genomic SARS-CoV-2 RNA fragments in site specific manner, and due to the presence of Zn(II) in different concentrations in intracellular fluids, Zn(II) represents an appealing cofactor for artificial

nucleases, while free Cu ions are almost not available in cellular environment.<sup>[150, 176-178]</sup> The incorporation of LNA (locked nucleic acids) are supposed to further improve stability and binding properties<sup>[146]</sup> and thus studies, including thermal melting studies with the dually labelled ORNs (ORN3 and ORN6), towards the use of Zn(II)-binding ligands on LNA modified ODN4 and ODN5, targeting stem-loop 4 (SL4) are now underway in the Kellett laboratories.

## 4.4 Investigations towards Copper-Free Click Chemistry for Optimized mRNA Vaccine Development.

### 4.4.1 Quantification of 3'-N<sub>3</sub> Labelling and SPAAC with DBCO-Trimannose-Moiety for baseclick's mRNA Vaccine Development

#### 4.4.1.1 Prologue

In response to the SARS-CoV-2 pandemic, the development of mRNA vaccines has garnered significant attention and as of December 2022, nine mRNA-based vaccines have been approved in at least one country. The most prominent and extensively administered vaccine candidates of Moderna (Spikevax) and Pfizer/BioNTech (Comirnaty) both encode for the spike gene of SARS-CoV-2 in order to induce an immune response against the virus.<sup>[179-180]</sup> As already mentioned in section 4.3.1, the highly mutating nature of the spike protein, enables immune evasion and also monoclonal antibody resistance.<sup>[122, 141, 143]</sup> Moreover, these vaccine candidates use lipid nanoparticles (LNPs) as protection and delivery system, which is associated with challenges, in terms of preparation, storage, inflammatory side effects and their accumulation in liver tissue.<sup>[181-183]</sup>

To overcome these challenges, the baseclick GmbH developed a novel mRNA-based SARS-CoV-2 vaccine candidate<sup>[184]</sup>, based on an IVT mRNA encoding for the SARS-CoV-2 N protein, which is supposed to be relatively conserved and exhibit a high immunogenicity.<sup>[185]</sup> This vaccine candidate, BCV-193N, does not only differ from the previous mentioned mRNA-based vaccines with regards to its encoded sequence, but also in its delivery approach. In contrast to the widely used LNP delivery system, BCV-193N uses a mannose-based targeting ligand, for targeted delivery to dendritic cells (DCs). The principle of this delivery method is based on the interaction of the sugar based targeting agent with C-type lectin (CTLs) receptors on DCs, inducing a receptor mediated endocytosis.<sup>[186]</sup> A similar approach has previously shown promising results for various applications, such as for targeting DCs with trimannosylated liposomes, mannosylated LNPs, or for targeted delivery of siRNA.<sup>[187-189]</sup>

BCV-193N vaccine candidate was prepared, starting with linearization of the circular plasmid DNA with the restriction enzyme BspQ1, followed by the IVT of a linear plasmid DNA, encoding the sequence the T7 promoter, the N protein of SARS-CoV-2 and a 120 nt poly(A) tail, *via* T7 RNA polymerase with the corresponding nucleotide triphosphates (NTPS). Apart from ATP,

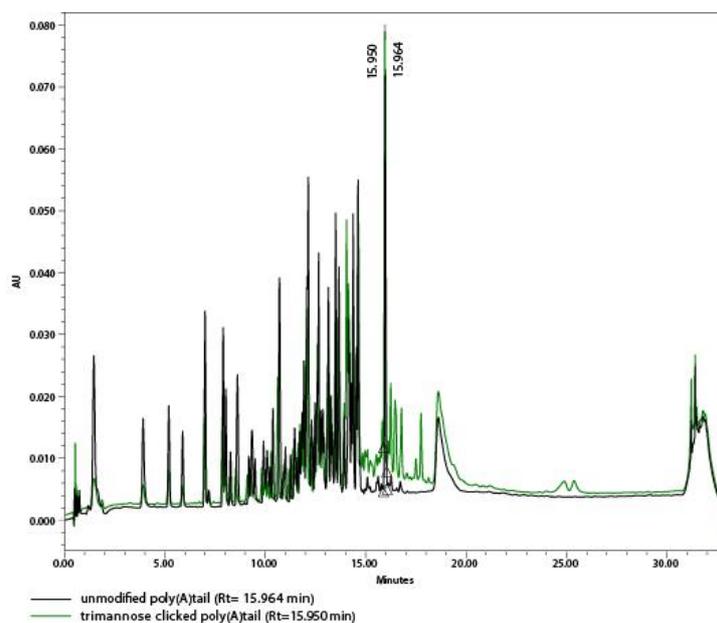


#### 4.4.1.2 Results and Discussion

In order to investigate the successful attachment of the sugar moiety to the N-protein encoding mRNA, eGFP mRNA (1222 nt), which sequence was designed by Croce *et al.*, was used as model system.<sup>[58, 78]</sup> Since we aimed to analyze the incorporation of the sugar moiety by RP-HPLC, and the size of the eGFP mRNA extended the accuracy level of our RP-HPLC, we wanted to cleave the poly(A)tail (120 nt), which should potentially bear A-N<sub>3</sub> and further the trimannose functionality, and analyze the smaller fragment *via* RP-HPLC. Our original idea was, to digest the eGFP mRNA with RNase A, a endoribonuclease that specifically degrades RNA after pyrimidine nucleotides (C and U), and thus should conserve the poly(A) tail. For a screening experiment unmodified eGFP mRNA (345-700 ng) was digested with RNase A (10 mg/ $\mu$ L), at 37°C for 5 or 30 min, using different amounts of enzyme (0.2-2  $\mu$ L). The samples were purified by EtOH precipitation and analyzed by analytical RP-HPLC, where the chromatograms indicated several peaks with a similar retention time. We assumed that these peaks are displaying mono-, di-, and trinucleotides, but since there was no peak with a significant longer retention time, which we expected for the poly(A)tail, the approach using RNase A for digestion was discarded.

Inspired by the work of Beverly and co-workers, we next wanted to digest the eGFP mRNA with RNase T1 (1000 U/ $\mu$ L), a endoribonuclease that specifically degrades ssRNA at G residues, again in order to retain the poly(A)tail.<sup>[190]</sup> Different conditions, with varying amounts of enzyme and incubation times (15-60 min) were tested and we tried to isolate the poly(A) tail subsequently with Oligo d(T)<sub>25</sub> magnetic beads or EtOH precipitation. When digesting, the eGFP mRNA (2.1-12.9  $\mu$ g) with RNase T1 (200-500 U) for 1 h at 37°C, followed by EtOH precipitation, HPL chromatogram indicated now a significant peak at 260 nm, with a higher retention time, which could potentially be assigned to the poly(A)tail. The same digestion was performed for the 3' -N<sub>3</sub> labelled and subsequent sugar clicked mRNA and the precipitated sample was analyzed by RP-HPLC. The unmodified and trimannose-clicked mRNA showed exactly the same pattern in the chromatogram (Figure 36), and since it was not possible to analyze the poly(A)tail-fragment by MALDI-TOF MS, because the device is only suitable up to 20 kDa (poly(A) tail + G+ Z, 122 nt,  $m/z$  [M+H]<sup>+</sup> calcd.: 38117.62), it was necessary to move on with another quantification method. Anyway MALDI-TOF MS was performed, however the

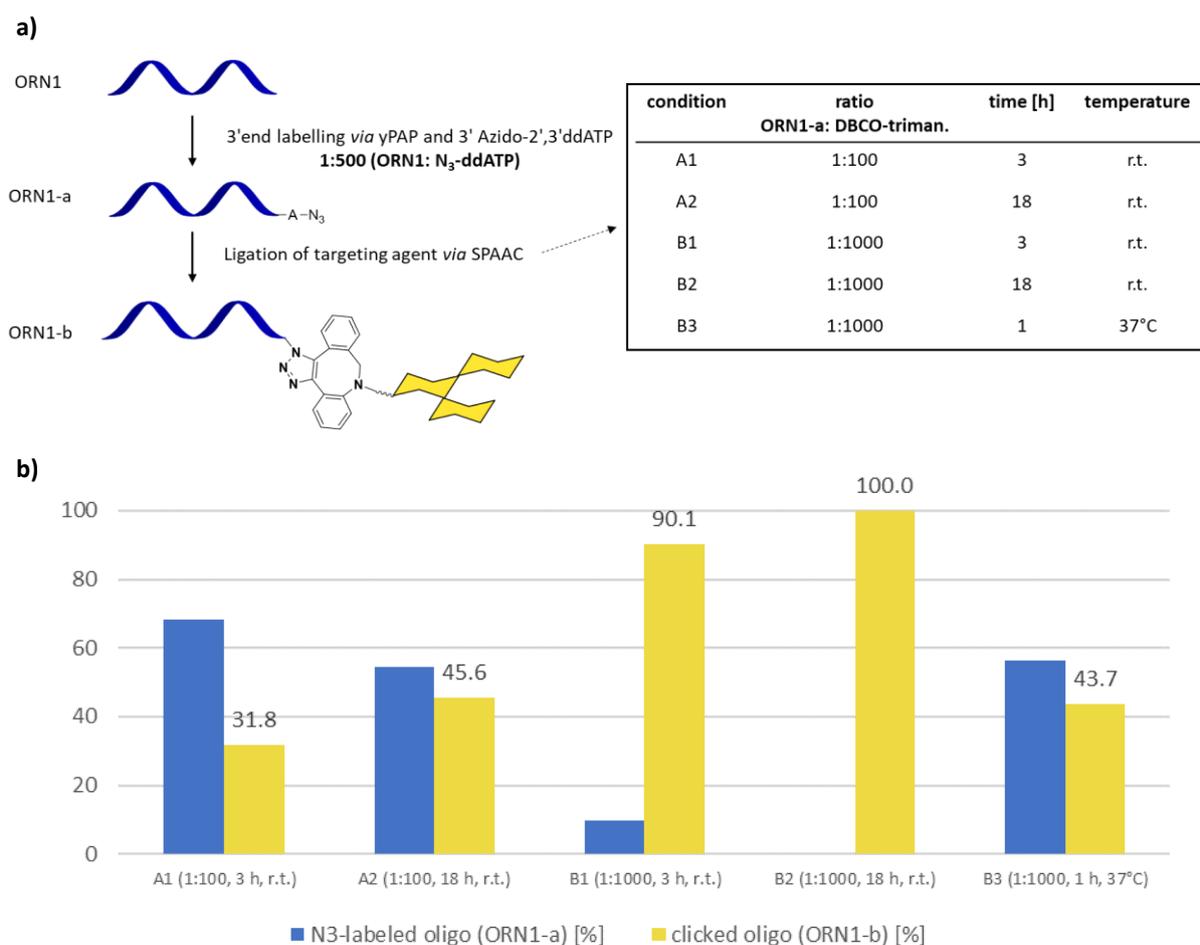
spectra indicated only a major peak at  $m/z$  17120, which cannot be assigned to the desired fragment.



**Figure 36.** Overlaid analytical RP-HPL chromatograms of isolated fragments of unmodified and trimannose clicked mRNA after digestion with RNase T1. The integrated peaks at 15.964 and 15.950 min were assumed to display the poly(A)tail of unmodified and trimannose clicked poly(A) tail fragments respectively.

The new approach is based on the digestion to single nucleotides and subsequent mass spectrometry analysis (LC-MS) of the cleaved fragments. By comparing the digestion products of unmodified, 3'-N<sub>3</sub> and 3'-trimannose clicked mRNA, we aimed to quantify the 3'-end labelling with 3'-azido-2',3'-ddATP and the subsequent SPAAC with the DBCO-bearing trimannose targeting ligand. In order to identify the best conditions for the individual steps and further quantification, we started with an oligoribonucleotide model system. Therefore, the individual steps were carried out on the unmodified 21 mer oligoribonucleotide (ORN1). The 3'-N<sub>3</sub> labelling was performed in reference to Croce *et al.*, in the presence of yPAP in a molar ratio of 1:500, ORN1 to 3'-azido-2',3'-ddATP respectively (2 h, 37°C). Next, different SPAAC conditions were tested (Figure 37), since previously the reaction has been performed using a 1:100 or 1:1000 ratio (3'-N<sub>3</sub> labelled mRNA: DBCO-trimannose), but no analysis was performed to justify the ratios or to prove their efficacy. The spin column purified samples were first analyzed qualitatively by MALDI-TOF and further quantitatively by LC-MS (Orbitrap). Although MALDI-TOF analysis was only used for qualitative analysis, it could be observed that for the reactions with a ratio of 1:100, the peak intensity of the 3'-N<sub>3</sub>-labelled ORN1 (ORN1-a) was almost as strong as the clicked ORN1 (ORN1-b), while for the ratio of 1:1000, the starting material ORN1-a was not observed anymore. In the LC-MS analysis, the UV signals were too

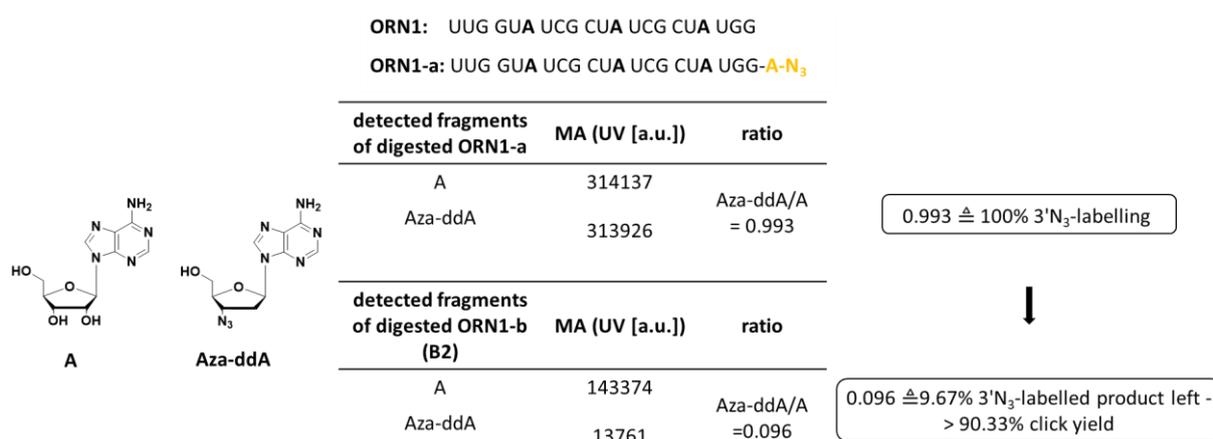
low to use them for quantification, thus we assumed all samples ionize to the same extent and used the mass trace for quantification. The generated data indicated, that the 3'-end labelling yields 92.2% of ORN1-a. For the SPAAC with a 1:1000 ratio after 3 h reaction time, only 9.9% of ORN1-a remained in the sample, while after 18 h, no ORN1-a could be detected any more (NSF= no signal found), which can either be caused by a weak signal, or there is indeed no ORN1-a left and the SPAAC led to completion. Again, the 1:100 ratio yielded after 18 h reaction time only 45.6% ORN1-b and condition B3 also only yielded 43.7% (Figure 37).



**Figure 37.** **a)** left) schematic representation of the ORN1 model system right) varying conditions for SPAAC between ORN1-a and DBCO-trimannose. **b)** SPAAC quantification by LC-MS (Orbitrap).

One sample set of unmodified ORN1, 3'-N<sub>3</sub>-labelled ORN1 (ORN1-a) and clicked ORN1 (ORN1-b), whereas the SPAAC was performed using conditions B2 (1:1000, r.t., 18 h) was further analyzed *via* analytical RP-HPLC confirming that the click was quantitatively as no ORN1-a could be detected any more (section 6.5.1, Table 11).

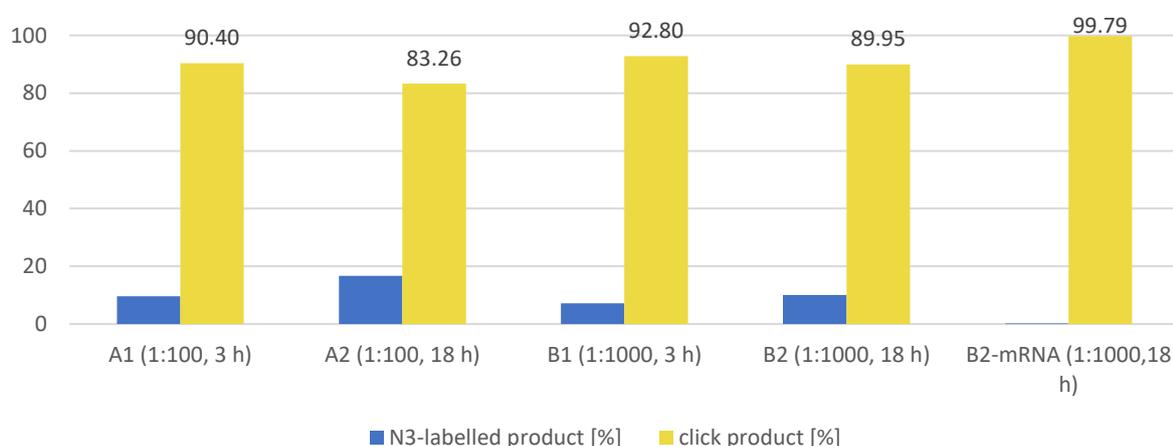
Since the yield of clicked product, analyzed by LC-MS, was significantly higher with the 1:1000 ratio, we now wanted to transfer the conditions to the N-protein encoding mRNA and quantify the SPAAC after digesting to single nucleotides. In order to compare the data from the ORN1 model system with the 1734 nt long mRNA, also samples from the 21mer model system were digested to single nucleosides *via* nucleotide digestion mix from NEB. While for the model system the products of the SPAAC conditions A1, A2, B1 and B2, were analyzed, for the N-protein encoding mRNA the SPAAC was only performed with the condition B2 (1:1000, 18 h, r.t.). ORN1, ORN1-a, ORN1-b and N-mRNA (mRNA), 3'-end labelled N-mRNA (mRNA-a) and clicked N-mRNA (mRNA-b) were digested and the cleavage products were subsequently analyzed *via* LC-MS. For reference, 3'-azido-2',3'-ddATP (Aza-ddATP) and 3'-azido-2',3'-ddATP clicked to the DBCO-trimannose were also digested. This time, the quantification for the model system was performed using the UV trace. In the digested sample of ORN1-a, the ratio of Aza-ddA/A was 0.096, indicating that the absorption of 3'-azido-2',3'-ddA (Aza-ddA) is roughly 3 times higher than A, when taking into account that ORN1 contains three As and assuming the 3'-end labelling and digest proceeded quantitatively. This ratio was then used to determine the amount of ORN1-a left in each sample of the individual SPAAC reactions, which further enables the yield calculation of the SPAAC. One example, how the SPAAC conversion yield was calculated, is given in Figure 38.



**Figure 38.** Structures of the digestion fragments A and Aza-ddA, used for the quantification of the SPAAC and exemplary calculation of the SPAAC yield for condition B2. MA= mass area.

For the mRNA samples, the UV signals were too low, thus the mass trace was used to determine the yields, where it was again assumed homogeneous ionization and that the azide elongation was quantitative. The results, displayed in Figure 39, indicate that all click reactions yield a

minimum of 83% (ORN1, condition A2). Comparing the same reaction ratios of A and B, it can be noticed that the yield of the SPAAC is higher after 3 h compare to 18 h reaction time. This observation cannot be explained in a straightforward manner, since the SPAAC is irreversible and the reaction is fairly slow.<sup>[27]</sup> One reason could therefore be that the samples were mixed up or the sensitivity of this quantification is not high enough. Either way, the click reactions with a ratio of 1:1000 yielded minimum 90.3%, including the mRNA sample (Figure 39). Taking into account that the azide elongation yielded 92.2% 3'-N<sub>3</sub>-labelled product, and the SPAAC with conditions B gave minimum 90% clicked product, the overall yield of the trimannose labelling is 83.0 %.



**Figure 39.** SPAAC quantification by digestion to single nucleosides. For determination of the click yield either the UV trace (A1, A2, B1 and B2) or the mass trace (B2-mRNA) was used.

Due to device issues, it was not possible to use the prepared reference material for quantification. Since several approximations were taken into account and for the mRNA SPAAC quantification, the mass trace was used, while for the model system the UV trace was applied, the generated data should be considered cautiously. For precise determination of the SPAAC yield on the N-mRNA, the experiments would need to be repeated, whereas more material would be needed in order to get sufficient signals for the UV trace. Further it would be necessary to prepare a calibration curve for the 3'-azido-2',3'-ddA and 3'DBCO-trimannose-clicked-2',3'-ddA fragments to guarantee reliable quantification.

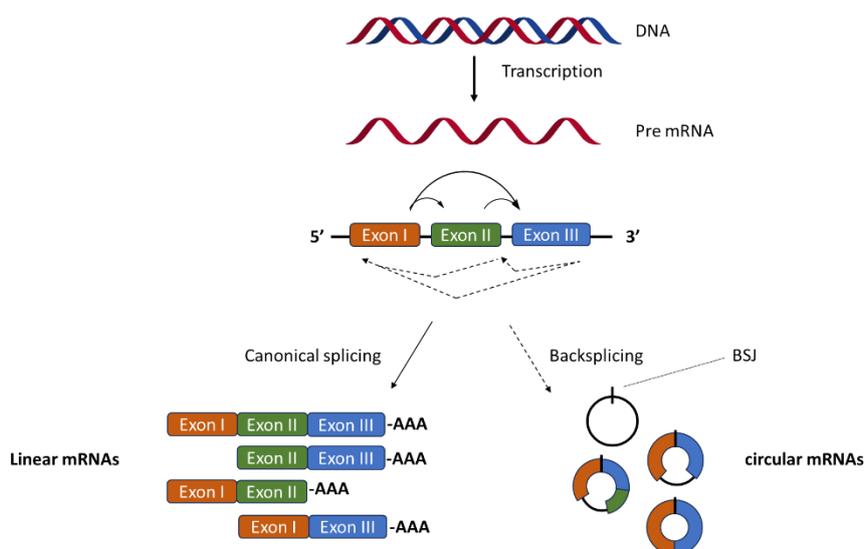
In summary, some preliminary, but nevertheless encouraging results have been obtained towards the quantification of a two-step 3'- end labelling of therapeutic mRNA with a sugar-based targeting agent. While our current LC-MS data indicates at least 82% labelled product, we envision that after a certain degree of optimization, the yield of the two-step labelling could be monitored and determined with higher precision, which would allow further improvements of the individual labelling steps.

## 4.4.2 Circularization of an Oligoribonucleotide *via* SPAAC for Potential Circular mRNA Application

### 4.4.2.1 Prologue

During the rapid development of mRNA-based vaccines, the circularization of coding RNAs into circular RNAs (circRNAs) attained substantial interest, as a way to achieve enhanced protein production. <sup>[191]</sup>

CircRNAs are a class of natural occurring coding or non-coding closed RNAs, which are generated in eukaryotic cells through back-splicing (Figure 40) <sup>[192]</sup> Due to their lack of free ends, circRNAs are not affected by exonuclease-mediated degradation, resulting in at least 2.5 fold longer life time, compared to linear RNAs. Thus, synthetic protein-coding circRNAs have garnered increased attention for numerous biological applications. Due to their lack of a m<sup>7</sup>G cap, circRNAs exhibit a different, cap-independent translation initiation process compared to their linear counterparts. One possibility therefore is the incorporation of an internal ribosome entry site (IRES), or a m<sup>6</sup>A modification upstream of the ORF. <sup>[191, 193-195]</sup> Apart from their higher stability, accompanied with enhanced protein production, Wesselhoeft and co-workers showed, that unmodified circRNAs are less immunogenic than unmodified linear mRNAs due to their ability to circumvent toll like receptor (TLR) sensing and further showed that circRNA are compatible with LNP-mediated delivery. <sup>[196]</sup> Thus, within the last years, the scope of circRNA applications has broadened, including the development of a circRNA-based SARS-CoV-2 vaccine. <sup>[194]</sup> For long RNA circularization, a ribozymatic approach, utilizing a permuted group I intron-exon (PIE) splicing method, optimized by Wesselhoeft *et al.*, has proven most applicable. <sup>[193]</sup> Since this method requires additional sequences apart from the IRES and the protein coding sequence, we were eager to facilitate circularization of coding RNA by the use of click chemistry. The idea of using click chemistry for circularization of nucleic acids has previously been introduced, however to the best of our knowledge only for DNA or synthetic 5'-N<sub>3</sub> and 3'-propargyl decorated oligoribonucleotide by Cu(I) catalyzed alkyne azide cycloaddition. <sup>[197-198]</sup> Thus, with our approach we wanted to enable circularization of RNA *via* SPAAC in order to avoid potential oxidative damage of RNA, which is potentially caused by the copper catalyst. <sup>[199]</sup>

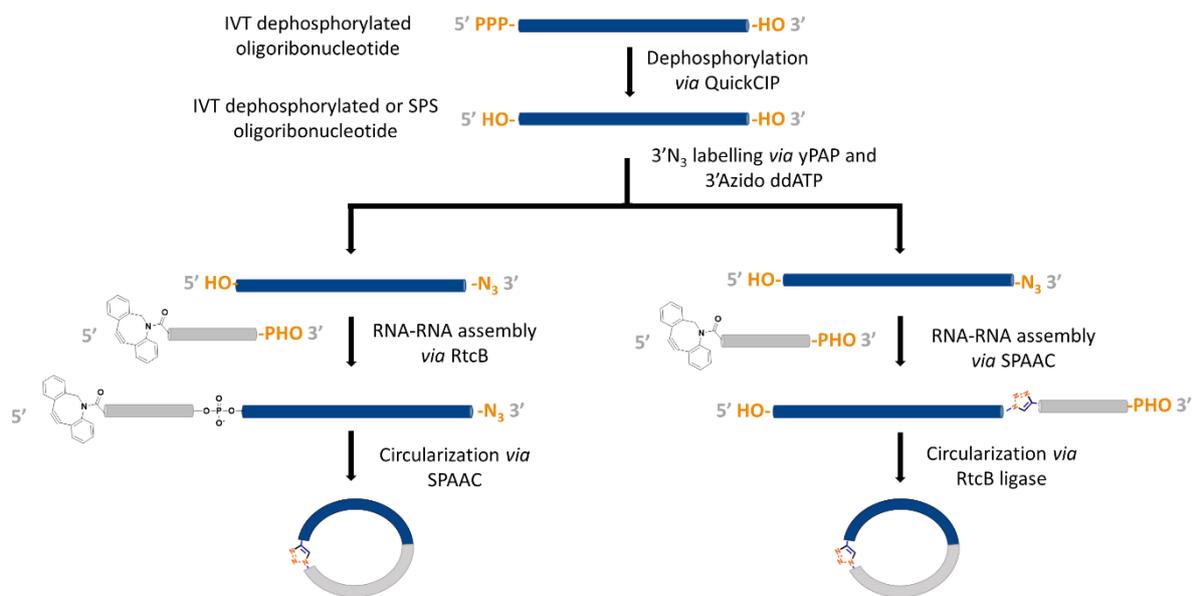


**Figure 40.** Schematic representation of linear and circular mRNA formation. Linear mRNAs are formed by canonical splicing whereby introns are eliminated from the pre-mRNA. Circular mRNAs are formed by backsplicing, where a covalent linkage between the downstream 3'-end of a pre-mRNA sequence and upstream 5'-end of the pre-mRNA strand is formed, leading to a backspliced junction (BSJ). The figure was prepared in reference to Curry-Hyde *et al.*<sup>[200]</sup>

#### 4.4.2.2 Results and Discussion

We aimed to investigate a universally applicable protocol for attaching an RNA sequence, which could potentially encode the IRES sequence or a modified linker oligoribonucleotide attached to a targeting agent, to the protein coding sequence of interest *via* Cu-free click chemistry, followed by cyclization of the generated RNA-RNA assembly *via* ligation. In the first step of our model system, the 5'-terminal triphosphate of the *in vitro* transcribed RNA oligonucleotide would be dephosphorylated to generate a 5'-OH end. Next, the 3'-end of the IVT mRNA will be labelled with an N<sub>3</sub>- functionality by incorporation of 3' Azido-2',3'ddATP *via* yPAP, followed by the ligation of a 5'-DBCO and 3'-phosphorylated (PHO) oligoribonucleotide *via* RtcB ligase, to its 5'-end. The generated RNA-RNA assembly, displaying 5'-DBCO and 3'-azido ends, should finally be circularized *via* SPAAC, after which, as a proof of concept, the final product will be treated with RNase R. The click reaction and chemical ligation could also be performed in the opposite order, thus apart from investigating the feasibility of the proposed circularization approach itself, we wanted to analyze what impact the order of SPAAC and ligation has on the overall strategy and if the SPAAC and ligation could even be achieved simultaneously (Figure 41). Commonly, ligation-related circularization methods use T4 ligase, which enables the formation of a phosphodiester bond between the 5'-phosphate

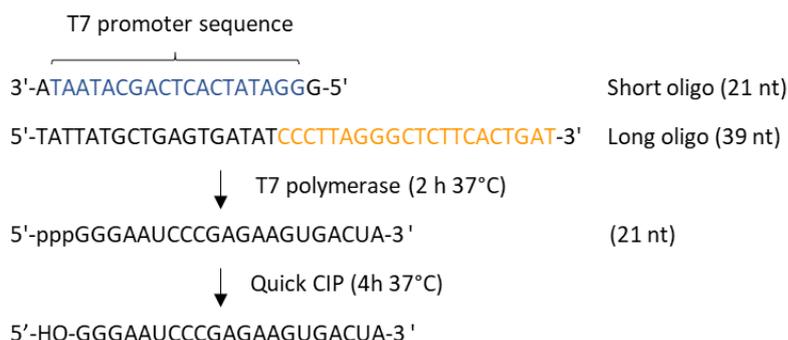
and the 3'-OH of RNAs.<sup>[201]</sup> Since we wanted to achieve simultaneous ligation and click, we designed a method using RtcB ligase, in order to avoid cross-circularization of the 5'-phosphate - 3'-OH IVT RNA without incorporation of the second, DBCO-bearing RNA, which could be caused if the 3'end labelling is not quantitatively in combination with the use of T4 ligase.



**Figure 41.** Schematic RNA circularization workflow. **a)** RNA-RNA assembly prepared by ligation via RtcB followed by circularization via SPAAC. **b)** RNA-RNA assembly prepared via SPAAC followed by circularization via RtcB.

Upon carrying out the circularization of our RNA-RNA oligonucleotide model system successfully, we aim to incorporate a second internal functionality to the 5'-DBCO- bearing oligoribonucleotide, to enable further modification with any molecule of interest, such as a targeting agent and transfer the model system to longer RNAs.

Initially, we wanted to start our approach with a IVT oligoribonucleotide. Therefore, a dsDNA template was designed, which was composed of a short, 21 mer oligonucleotide displaying the T7 promoter sequence, and a longer 39 mer oligonucleotide. Upon annealing of the two oligonucleotides and IVT via T7 polymerase (2 h, 37°C), we wanted to gain a 21 mer oligoribonucleotide, which, after digestion of the template oligos via DNase I (1 h, 37°C) should subsequently be dephosphorylated via Quick CIP (4 h, 37°C) as displayed in Figure 42.

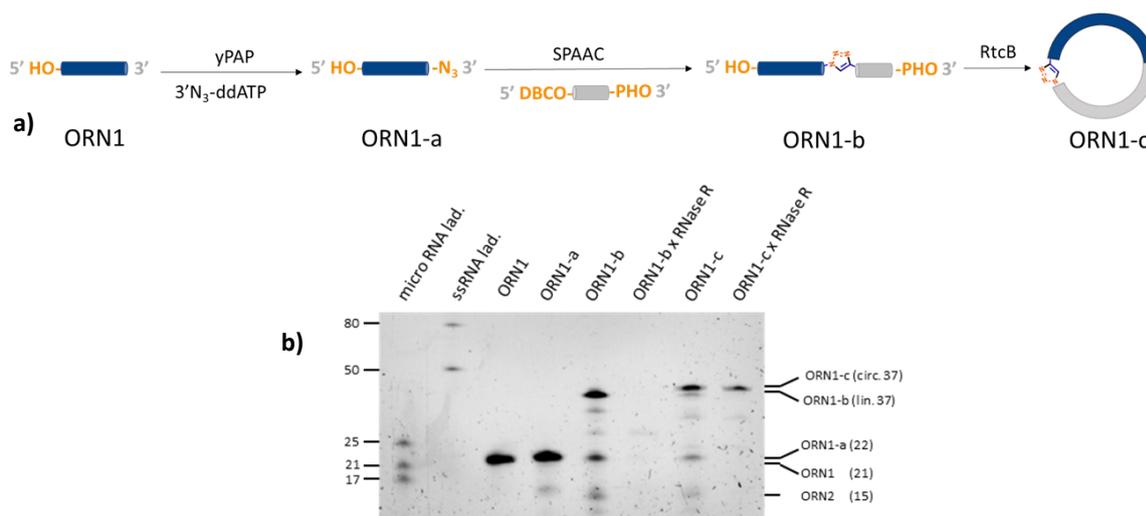


**Figure 42.** IVT approach of 21 mer oligoribonucleotide.

After the IVT and digestion of the template DNA, the spin column purified IVT oligoribonucleotide was analyzed *via* MALDI-TOF-MS, indicating a significant mass peak of  $m/z$  7041.39 ( $[M+H]^+$ ), which can be assigned to the desired 21 mer ( $[M+H]^+$   $m/z$  calcd.: 7037.90) oligoribonucleotide. However, the spectrum also indicated major impurities. Thus, to investigate whether these impurities are coming from a partial degradation of the template DNAs, analytical RP-HPLC was performed. The chromatogram indicated only one peak per template oligonucleotide, confirming that the template DNA was not degraded. Another possibility for the impurities would be insufficient digestion of the template DNA by DNase I. Thus, the digestion was repeated and a control experiment was performed to investigate whether the DNase I was functioning properly. Since the results indicated that the impurities were not caused by insufficient digestion of the dsDNA template, we assume, that the design of the dsDNA template itself caused unspecific transcription of oligoribonucleotides. The template DNA was only partly double stranded and rather short, therefore the T7 polymerase might not be able to fully attach to the template or display a low affinity towards the template, yielding different RNA oligonucleotides with varying lengths, which caused the major impurities in the MALDI-TOF-MS spectrum. Thus, it would have been necessary to design a longer, dsDNA template. Since we wanted to proceed with the actual approach for circularization, we decided to continue the experiments with a solid phase synthesized 21 mer oligoribonucleotide ORN1 (Figure 43a)

The first step was the incorporation of a 3'-N<sub>3</sub> functionality, which was performed *via* γPAP and 3'-azido-2', 3'-ddATP (2 h, 37°C). After spin column purification, the sample was analyzed *via* MALDI-TOF-MS, indicating a significant peak at  $m/z$  7004.92 ( $[M+H]^+$ ), which can be assigned to the desired 3'-A-N<sub>3</sub> labelled ORN1, ORN1-a ( $[M+H]^+$   $m/z$  calcd.: 7005.20). Additionally, dPAGE (6.1 A) was performed, where the image indicated a faint smaller

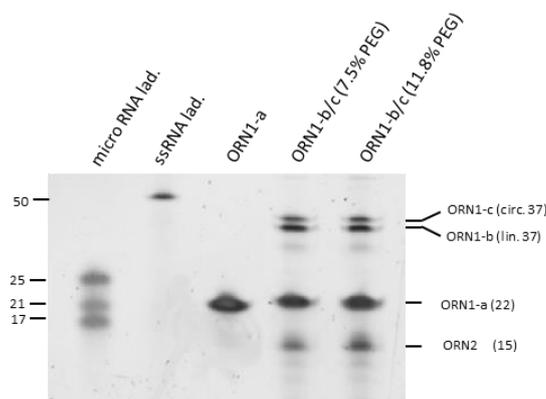
molecular weight impurity, which could be a side or degradation product from the enzymatic reaction. Since the impurity could not be assigned to a significant peak in the MS spectrum, we continued with the next step, which would either be the SPAAC with the 5'-DBCO end of the 15 mer oligonucleotide or the ligation of *via* RtcB ligase to connect the 3'-PHO end of the 15 mer (Figure 41). Since we first needed to investigate appropriate RNA-RNA click conditions, we started with the SPAAC (Figure 41a). For the ligation of two RNA strands, an appropriate salt concentration is crucial in order to bring the two negatively charged strands together. Thus, different conditions were investigated, whereas a modified protocol from and Taemaitree co-workers gave the desired RNA-RNA oligo (37 mer) assembly, ORN1-b, as confirmed by dPAGE (Figure 43b) .<sup>[202]</sup> The gel image further indicates some impurities, which can be attributed to the formed side products of the lower molecular weight impurity from the azide elongation step. The successful SPAAC was carried out at r.t. overnight, in a 1:2.5 ratio of ORN1-A-N<sub>3</sub> and 5'DBCO oligonucleotide (ORN2,15 mer) respectively. Next, we wanted to circularize the RNA-RNA assembly by ligating its 5'-OH end with the 3'-PHO end *via* RtcB ligase, which was performed for 2 h at 37°C. After spin column purification the sample was again analyzed by dPAGE, featuring a new band which is running slightly slower than the linear, clicked 37 mer, ORN1-b (Figure 43b). To investigate whether this new band can be attributed to the circ-RNA assembly, the sample was treated with RNase R, a 3'-5' exoribonuclease. Two digestion reactions were set up, whereas one was containing the sample from the circularization step (ORN1-c) and the other one, a control sample, was containing the linear SPAAC click product (ORN1-b). Both samples were treated equally (incubation for 2 h at 37°C) and analyzed by dPAGE after spin column purification. Indeed, while the ORN1-b has been completely digested by RNase R, the 37mer which was attributed to the circularized product ORN1-c, was still visible, indicating successful circularization (Figure 43b).



**Figure 43. a)** Reaction scheme of the RNA-RNA circularization approach **b)** dPAGE (15%) gel image after staining with GelRed® Nucleic Acid Stain.

Next, we wanted to perform the enzymatic ligation previous to the SPAAC, to investigate whether this reaction order would have an impact on the yield of the circularization product or if the ligation and SPAAC can be even performed simultaneously (Figure 41b). Therefore, ORN1-a was incubated with 5'-DBCO-3'-PHO ORN2 and RtcB ligase using the same conditions as for the previous ligation experiment. After purification the sample was analyzed by dPAGE, indicating the bands of the 22mer (ORN1-a), 15mer (ORN2) and the desired 37mer, ORN1-b (data not shown). The band of the ligated RNA product was however barely visible, suggesting that the ligation was low yielding. Therefore, we set up another ligation experiment, where we added PEG6000 as molecular crowding agent, which could potentially increase the yield of the ligation.<sup>[203]</sup> Two reactions with either 7.5% or 11.8% of PEG6000 in the overall sample were performed, where the two oligos ORN1-a and ORN2 were added in equal amounts and incubated again for 2 h at 37°C. After purification the reaction products were analyzed by dPAGE (Figure 44). The gel image indicated that the ligation and SPAAC have occurred simultaneously, since apart from the 22 mer (ORN1-a), the 15 mer (ORN2) and the expected linear 37 mer (ORN1-b), an additional band, running slightly slower than the linear 37 mer is visible. Since the migration time of this additional band is equal to the previous circularized 37 mer (ORN1-c), we assume that indeed the ligation and SPAAC have taken place at the same time (ORN1-b/c, Figure 44). The yield of the circularized product (ORN1-c) seems to be lower compared to the first approach where the SPAAC was performed prior to the ligation. However, this could potentially be improved by a longer reaction time, since while the first

SPAAC was performed over night at r.t (ca. 18 h), the second SPAAC was performed for only 2 h at 37°C. Further, there is no significant difference between 7.5 and 11.8% visible, thus apparently already 7.5% of PEG6000 enables that enzymatic ligation and SPAAC occur simultaneously.



**Figure 44.** dPAGE (15%) gel image after staining with GelRed® Nucleic Acid Stain, displaying the bands of the simultaneous SPAAC and enzymatic ligation in the presence of PEG6000 (lane 4 and 5).

In summary our preliminary results showed, that enzymatic ligation and SPAAC of two small RNA strands can be performed simultaneously. Further improvements would be necessary to obtain higher yields of the circularized RNA-RNA assembly, but we already enabled a one-pot circularization within only 2 h. Nevertheless, this approach was not pursued any further, since our established method of labelling the 5'-end and the 3'-end of nucleic acids with a clickable functionality, followed by circularization, is more favorable compared to the approach to achieve circularization *via* click reaction and enzymatic ligation.<sup>[87]</sup>

## 5. Thesis Discussion

### **Click Chemistry Enables Rapid Amplification of Full-Length Reverse Transcripts for Long-Read Third Generation Sequencing**

In order to prepare sequencing libraries covering full-length reverse transcripts, the Smart-seq2 protocol, involving a template switching (TS) reaction, is undoubtedly the leading technology on the market. While this method enables groundbreaking achievements, template switching brings forth various issues, such as inefficient ligation, an overrepresentation of sequences containing 5'-guanosine, and the occurrence of significant artifacts and erroneous alternative transcripts. With the work, published in this thesis, we present an innovative approach using click chemistry for preparing long-read sequencing libraries from polyadenylated RNA. The method involves 3'-azido-labelling of reverse transcribed cDNA through an anchored poly(dT) primer, followed by CuAAC and subsequent PCR amplification to produce click chemistry-based dsDNA libraries for Third Generation Sequencing. Each step in this novel library preparation is reproducible, high yielding, and straightforward to execute. Herein, the incorporation of an extra, quadruple degenerated nucleotide at the 3'-end of the forward primer significantly enhances PCR quality, enabling polymerase-mediated amplification of long triazole-containing ssDNA templates.

In Nanopore sequencing experiments, directly comparing click chemistry-based and standard protocol libraries (template switching), our data shows competitive performance in both amplicon and transcriptome sequencing. The unbiased addition of sequencing adapters simplifies RNA-seq protocols, making it suitable for methods like SHAPE-seq or sequencing of RNA modifications. (e.g., m<sup>1</sup>A or pseudo uridine).<sup>[204]</sup>

In conclusion, our innovative library preparation method ensures reliable generation and amplification of full-length reverse transcripts for long-read sequencing, avoiding templated ligation, artifactual chimeras, and over-representation of 5'-guanosine-containing sequences. Compared to PAC-seq, transcripts longer than 1 kb are now accessible. Thus, our method serves as an attractive alternative to Smart-Seq2, potentially laying the groundwork for more advanced workflows in the future.<sup>[70]</sup>

## Orthogonal End Labelling of Oligonucleotides through Dual Incorporation of Click-Reactive NTP Analogues

In recent decades, advancements in site-specific bioconjugation of nucleic acids have significantly expanded the scope of nucleic acid technology. However, while numerous chemical and enzymatic options exist for modifying the 3'-ends of DNA and RNA molecules after synthesis, the 5'-end labelling is often limited. The work, published in this thesis, introduces a versatile chemoenzymatic procedure based on click chemistry to modify the 5'-end of single-stranded nucleic acids.

Originally, phosphoramidates were synthesized with the prospect of serving as T4 PNK substrates, as reported by Anthony and co-workers.<sup>[97]</sup> However, considering the potential susceptibility of phosphoramidates like **2** or **3**<sup>[87]</sup> to acid-mediated solvolysis,<sup>[98]</sup> we extended our efforts to create a series of phosphate monoester-linked alkynes (compounds **4–6**) and azide (compound **7**).<sup>[87]</sup> These ATP analogues were obtained with yields of up to 68% through nucleophilic substitution, employing either the corresponding alkyl halide or *p*-toluenesulfonate. Additionally, organothiophosphate **8** was synthesized using similar conditions and compounds **9** and **10** were purchased for subsequent reactivity screening purposes.

With these compounds, we proceeded to examine their suitability for labelling the 5'-OH of oligonucleotides, as a model system, *via* T4 PNK catalyzed reaction. Following multiple improvements to the labelling methodology, we successfully accomplished the transfer of the click-functionalized  $\gamma$ -phosphate to the 5'-OH of the oligonucleotide model system (17mer and 43mer oligonucleotide) through T4 PNK mediated catalysis, as confirmed by analytical RP-HPLC and by subsequent click reactions with fluorescent dyes. Notably, only phosphoesters were suitable as T4 PNK substrates, where shorter linkers had superior labelling efficiency compared to longer linkers. With the best performing ATP analogues, alkyne **4** and azide **9**, successful labelling on DNA as well as RNA oligonucleotides were achieved followed by CuAAC or SPAAC click reactions with fluorescent dyes or click modified oligonucleotides. Remarkably our approach allows for compatibility with a 3'-labelling reaction catalyzed by TdT, which further allowed us to generate a multicomponent chimeric DNA/DNA assembly and circularization of a dually click-modified (3'-N<sub>3</sub>, 5'-alkyne) ssDNA 100-mer *via* SPAAC.

Further to the work, represented in the publication, the synthesis of DBCO and allyl-functionalized ATP analogues was performed, in order to further broaden the scope of subsequent click reaction, upon successful labelling via T4 PNK. Although the chemoenzymatic 5'-labelling could not be achieved with these compounds, they could be of interest for other applications such as Third Generation Sequencing.

In summary, our novel chemoenzymatic method for labelling DNA and RNA at the 5'-end of with alkynes or azides followed by click reactions with oligonucleotides as well as small-molecules, pave the way for novel synthesis approaches for complex oligonucleotide structures and procedures essential for next-generation sequencing and the advancement of new nucleic acid therapeutics.

### **Click Modified Cu(II)-Dependent Nucleic Acid-Based Artificial Ribonuclease for Targeted cleavage of SARS-CoV-2 Genomic RNA**

In this chapter, encouraged by promising ASO therapeutics, targeting a variety of infectious, human diseases,<sup>[141]</sup> we aimed to develop an novel antisense oligonucleotide related therapeutic to inhibit viral SARS-CoV-2 replication.

Herein, a chemically modified oligonucleotide bearing a Cu-binding ligand, targeting a highly conserved stem loop sequence of the SARS-CoV-2 genomic RNA, should, upon binding, form a DNA-RNA hybrid. After complexation of copper and reduction of Cu(II) to Cu(I), the thereby formed ROS should preferably lead to site-specific cleavage of the SARS-CoV-2 RNA and further hinder viral replication. After identification a potential target sequence, namely the SL4 and off-target SM2, we designed RNA oligonucleotides mimicking the target and off-target sequences stem loop respectively. Further, 10 oligonucleotide sequences (ODN1-ODN6), with lengths between 19 and 31 nucleotides, were designed to target the SL4, covering different positions of the hairpin structure. To enable the attachment of a Cu-binding ligand through click chemistry, one C8-alkyne modified base was incorporated in each strand, where, in order to investigate optimal binding affinity and target specificity, different positions for the internal C8-alkyne modification were chosen. These modified oligonucleotides were then ligated to

the Cu-binding ligand "DC-Py-N<sub>3</sub>" (provided by the Kellett laboratories) *via* CuAAC click reaction.

The obtained C8-alkyne and Cu-ligand modified oligonucleotides were next analyzed towards their target and off-target affinity. Therefore, binding affinity analysis *via* native PAGE and microscale thermophoresis (MST) were performed. Although we face limitations of the selected method, the MST results matched with the binding analysis *via* nPAGE, revealing ODN4\* and ODN5\* as the most promising oligonucleotide candidates for the desired approach.

Overall, some preliminary, but nevertheless encouraging results have been obtained towards the novel approach for targeted, oligonucleotide based, cleavage of the SARS-CoV-2 genomic RNA, which optimizations are currently investigated at the Kellett laboratories.

### **Investigations towards Copper-Free Click Chemistry for Optimized mRNA Vaccine Development**

This part of the thesis has been aimed investigate and improve copper-free, stain promoted azide-alkyne cycloaddition for mRNA vaccine development. In one aspect we investigated the efficacy of the 3'-N<sub>3</sub> labelling and following SPAAC with DBCO-trimannose-moiety for baseclick's mRNA vaccine development. To examine the effective ligation of the targeting ligand to the mRNA encoding the N-protein, several approaches have been investigated. Our initial idea, to digest our model system, namely eGFP encoding mRNA with RNase A. We anticipated that this endoribonuclease, that specifically degrades RNA after pyrimidine nucleotides (C and U), should conserve the poly(A) tail, with or without click modification. However, due to unspecific cleavage pattern as confirmed by analytical RP-HPLC, this approach had to be discarded.

Next, we utilized RNase T1, an endoribonuclease that specifically degrades ssRNA at G residues, again in order to retain the poly(A)tail. Although this approach allowed to isolate the poly(A) tail, the approach had to be discarded since the applied quantification and qualification methods (analytical RP-HPLC, MALDI-TOF MS) were not sufficient to distinguish between labelled and unmodified poly(A)tail.

With digestion to single nucleotides, we were finally able to successfully quantify the 3'-N<sub>3</sub> labelling and following SPAAC with DBCO-trimannose-moiety by LC-MS. Therefore, a model system, namely a 21 mer oligoribonucleotide, was utilized and different reaction conditions were quantified and the best conditions were transferred to the N-protein encoding mRNA. Although the investigated method still required optimization for precise data generation, we have obtained initial yet promising outcomes in the quantification of a two-step 3'-end labeling process for therapeutic mRNA using a sugar-based targeting agent. Our latest LC-MS data suggests a minimum of 82% labeled product. We expect that, with further improvements, the refinement of the monitoring and precision of the yield in the two-step 3'-mRNA labelling process can be achieved.

Additionally, we investigated the circularization of an RNA-RNA oligoribonucleotide assembly *via* SPAAC and enzymatic ligation for potential circular mRNA application. In the initial stage of our 21mer RNA model system, the 5'-OH of our RNA oligonucleotide was labelled with a N<sub>3</sub>-functionality by integrating 3' Azido-2',3'ddATP through  $\gamma$ PAP. This step was followed by the ligation of a 5'-DBCO and a 3'-phosphorylated (PHO) oligoribonucleotide *via* RtcB ligase at its 5'-end. With the resulting RNA-RNA assembly, featuring 5'-DBCO and 3'-azido ends, we successfully achieved circularization through SPAAC, as confirmed by treatment with RNase R. Notably, the click reaction and chemical ligation could be performed simultaneously.

Although enhancements are required to achieve higher yields in the circularized RNA-RNA assembly and transfer the model system to protein encoding mRNA, we have successfully enabled a one-pot circularization within 2 h.

## 6. Experimental Section of Unpublished Work

### 6.1. General Experimental Section and Materials

**Materials and methods.** Chemicals were purchased from Sigma-Aldrich/Merck, BroadPharm, ABCR, Acros Organics or VWR and were used without further purification. Solvents were of reagent grade or purchased in septum-sealed bottles stored under an inert atmosphere. Unless otherwise specified, all reactions were magnetically stirred under a positive pressure of Argon (Ar). Reactions and chromatography fractions were monitored by qualitative thin-layer chromatography (TLC) on silica gel F<sub>254</sub> TLC plates from Merck KGaA and visualized by UV illumination or monitored by analytical RP-HPLC. Flash column chromatography was performed on silica gel (40-63  $\mu\text{m}$ ) from Merck KGaA. Dialysis membrane Spectra/Por<sup>®</sup> Biotech CE MWCO 100–500 Da (10 mm surface width) was purchased from Repligen and used as received. The DC-Pyr-N<sub>3</sub> ligand was provided by Alex Gibney (Kellett Group, School of Chemical Sciences, Dublin City University).

**Nuclear magnetic resonance (NMR).** NMR spectra were recorded on a Bruker Avance Neo (500 MHz) or Bruker Avance III (800 MHz) spectrometers. <sup>1</sup>H NMR shifts were calibrated to their residual protons of the deuterated solvent: CD<sub>2</sub>HOD (3.31 ppm), HDO (4.79 ppm), CHCl<sub>3</sub> (7.26 ppm). <sup>13</sup>C NMR shifts were calibrated to the residual solvent: CD<sub>3</sub>OD (49.00 ppm), CDCl<sub>3</sub> (77.16 ppm). The chemical shifts ( $\delta$ ) are given in ppm, the coupling constants (J) in Hz. Multiples are abbreviated as follows: s= singlet, d= doublet, t= triplet, q= quartet, dd= doublet of doublets, ddt= doublet of doublet of triplets, m= multiplet. The numbering in the assignments does not follow IUPAC rules and neither residual solvent signals nor tris-triethylammonium salt signals (TEAA buffer from RP-HPLC purification: quartet around 2.90-3.20 ppm, singlet around 1.90 ppm and triplet around 1.20-1.30 ppm) have been assigned for clarity. All NMR spectra were analyzed using the software MestReNova 14.1.1 from Mestrelab Research S. L.

**Mass Spectrometry (MS).** High resolution mass spectra (HRMS (ESI)) were recorded by the analytical section of the Department of Chemistry of the Ludwig-Maximilians-Universität Munich on a LTQ FT-ICR spectrometer from Thermo Finnigan GmbH. Matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectra were recorded on an autoflex<sup>®</sup> maX. Before MALDI-TOF MS, the samples (2  $\mu\text{L}$ ) were desalted on a 0.025  $\mu\text{M}$  VSWP

filter (Millipore) and 1  $\mu\text{L}$  of the desalted sample was applied to a stainless-steel sample plate using dried droplet method with a 3-hydroxypicolinic acid matrix (1  $\mu\text{L}$ , HPA: 25 mg 3-hydroxypicolinic acid, 5 mg ammonium citrate, 5  $\mu\text{L}$  15-crown-5 in 0.5 mL  $\text{H}_2\text{O}/\text{MeCN} = 1:1$ ). The obtained spectra were analyzed with FlexAnalysis 3.4 from Bruker Daltonics and processed with Excel. The molecular weight of the oligonucleotides was calculated using the HELM monomer toolbar from ChemDraw (version 22.0.0).

**Reversed-Phase High-Performance Liquid Chromatography (RP-HPLC).** Semi-preparative RP-HPLC was performed on a 1260 Infinity II Manual Preparative LC System from Agilent (Agilent Infinity II 12260 VWD: G7114A) equipped with the column VP 250/10 Nucleodur 100-5 C18 ec from Macherey-Nagel. A flow rate of 5 mL/min with varying gradients between 0-10% and 0-60% of buffer B over 45 min was applied for purification. The following buffer system was used: buffer A: 100 mM  $\text{NEt}_3/\text{HOAc}$  (pH 7.0) in  $\text{H}_2\text{O}$  and buffer B: 100 mM in  $\text{NEt}_3/\text{HOAc}$  (pH 7.0) 80% (v/v) acetonitrile. Analytical RP-HPLC was performed on a Waters Alliance (2695 Separation Module, 2998 Photodiode Array Detector) equipped with a XBridge<sup>®</sup> Oligonucleotide BEH C18 Column (30  $\text{\AA}$ , 2.5  $\mu\text{m}$ , 2.1 mm x 50 mm) using a flow rate of 1.5 mL and the same buffer system as mentioned above. Unless otherwise specified, varying gradients, between 0-10% and 0-60% buffer B from 0  $\rightarrow$  8 min, followed by 10-85% up to 60-85% buffer B from 8  $\rightarrow$  10 min, were applied.

**Determination of ATP and oligonucleotide concentrations.** The absorbance of ATP tetrabutylammonium salt,  $\gamma$ -phosphate modified ATP analogues and synthesized oligonucleotides was measured at 260 nm using an Implen NanoPhotometer<sup>®</sup> P300. The extinction coefficients ( $\epsilon$ ) of the oligonucleotides were calculated using the OligoAnalyzer<sup>™</sup> Tool from Integrated DNA Technologies. For internal C8-alkyne modified bases, the internal "Int 5-Octadiynyl dU" (/i5OctdU/) modification of the OligoAnalyzer<sup>™</sup> Tool was applied for calculations. For ATP tetrabutylammonium salt and for ATP disodium salt derivatives, an assumed extinction coefficient of  $\epsilon = 15400 \text{ L mol}^{-1} \text{ cm}^{-1}$  was applied, as these modifications are not expected to substantially alter the absorption properties of unmodified ATP. The concentrations of the oligonucleotides and ATP were further calculated *via* Beer-Lambert law.

$$c = \frac{A}{\varepsilon \times d}$$

*A ... absorbance, c ... concentration  $\left[\frac{\text{mol}}{\text{l}}\right]$ ,*

*$\varepsilon$  ... molar extinction coefficient  $\left[\frac{\text{l}}{\text{mol} \times \text{cm}}\right]$ , d ... optical path length [cm]*

### **Polyacrylamide Gel Electrophoresis Analysis (PAGE).**

- A) Denaturing PAGE (dPAGE): Unmodified RP-HPLC purified oligonucleotides and spin column purified oligonucleotide products were separated by either 10 or 15% denaturing urea (7 M) polyacrylamide gel electrophoresis in Tris-Borate-EDTA (TBE) buffer (1X: 0.13 mM tris, pH 7.6, 45 mM boric acid, 2.5 mM EDTA) at constant voltage (150 or 200 V) for 40 min - 1.5 h. The gels were stained for 30 min using GelRed® Nucleic Acid Stain (Millipore) according to the manufacturers protocol (1X dilution in 1X TBE buffer). The bands were detected on Gel Doc™ EZ Imager (Bio-Rad) using a UV Sample Tray (#1708271MP from Bio-Rad) and analyzed with Image Lab™ software. Single-stranded DNA oligo length standard (10/60 ladder) was purchased from Integrated DNA Technologies, microRNA marker and Low Range ssRNA ladder were purchased from New England Biolabs. RNA loading dye (2X) was purchased from Thermo Fisher Scientific.
- B) Native PAGE (nPAGE): Binding affinity and cleavage analysis of C8-alkyne modified and ligand-clicked oligonucleotides were performed by 20% native polyacrylamide gel electrophoresis in Tris-Borate-EDTA (TBE) buffer (1X: 0.13 mM tris, pH 7.6, 45 mM boric acid, 2.5 mM EDTA) at constantly 70 V for 4 h. The bands were stained for 10 min using SYBR™ Gold (Invitrogen; 1X dilution in nuclease free H<sub>2</sub>O) and then imaged using different filters (Cy5, Cy3 and blue filter) on a Syngene™ G:BOX Mini 9- Gel Documentation System. Orange DNA Loading Dye (6X) was purchased from Thermo Fisher Scientific. For each gel a pre-run at constantly 70 V for 1 h was performed.

## 6.2. Click chemistry-based library preparation and direct cDNA sequencing

### Preparation of click-ligated first-strand cDNA

The eGFP mRNA and first-strand cDNA synthesis, as well as the azide elongation with 3'-Azido-2',3'-ddGTP and the subsequent CuAAC with alkyne adapter (AA1) were performed in reference to Schönegger *et al.*<sup>[78]</sup>

### Second Strand synthesis

The second strand synthesis, to generate a dsDNA library was prepared using a modified protocol in reference to the manufacturer's instructions (Oxford Nanopore Technologies; previously SQK-DCS109, relaunched as SQK-LSK114)<sup>[82]</sup>. Two replicates were performed, whereas for each experiment 16.7-17 µL of clicked-cDNA (34-58 ng/µL), 8 µL FP3 (10 µM, provided by baseclick GmbH), 25 µL of LongAmp Taq 2x Master Mix (New England Biolabs) and RNase free H<sub>2</sub>O, to reach a total volume of 50 µL, were added to a 200 µL PCR vials and mixed well. For the second strand synthesis the following profile was applied: 1 min at 94°C (denaturation), 1 min at 57°C (annealing) and 15 min at 65°C (elongation). Afterwards the dsDNA products were further purified using the QIAquick PCR Purification Kit from Qiagen and the concentration of the combined dsDNA sample was measured with the Qubit 1xdsDNA HS kit on a Qubit Fluorometer.

### ONT Sequencing

The purified dsDNA products were converted into sequencing libraries using the direct cDNA sequencing kit (SQK-DCS109) from Oxford Nanopore Technologies and subsequently sequenced for 24 h on a MinION flow cell (R10.3) with a MinION sequencing device (Oxford Nanopore Technologies).

### Bioinformatic Analysis

The signals were recorded by the MinKNOW software (version 21.02.1) and the basecalling, to generated FASTQ files from FAST5 files was performed by the MinKNOW integrated software Guppy (version 4.3.4). Mapping against the eGFP mRNA template (Schönegger *et al.*, Supporting information Figure S5) was performed using the NanoPipe software (<https://www.bioinformatics.uni-muenster.de/tools/nanopipe2/generate/index.pl?lang=en>)

<sup>[205]</sup>. The coverage plot was generated with SAMtools (<http://www.htslib.org/>)<sup>[206]</sup> and the normalized read number was assembled with Excel and Affinity designer (version 2.1.1.1847).

## 6.3 Experimental section of unpublished data towards the 2<sup>nd</sup> Publication

### 6.3.1 Chemical Synthesis

#### **General procedure I.**

Conversion of ATP disodium salt into ATP tetrabutylammonium salt. The synthesis of  $\gamma$ -O-linked ATP analogues *via* alkylation on the  $\gamma$ -phosphate, reported by Hacker *et al.*, affords the conversion of the starting material, ATP disodium salt, into its tetrabutylammonium salt to guarantee strictly anhydrous conditions.<sup>[98, 100]</sup> Hence, 4 g Chelex<sup>®</sup> 100 sodium form were dissolved in 20 mL H<sub>2</sub>O, transferred into a glass column and washed with 500 mL H<sub>2</sub>O. Afterwards, 10 g tetrabutylammonium bromide were dissolved in 100 mL H<sub>2</sub>O, flushed through the column and washed with 500 mL H<sub>2</sub>O. Then, 300-550  $\mu$ mol of ATP disodium salt were diluted in 10 mL H<sub>2</sub>O, applied to the column and afterwards fraction collection was started. The column was washed with 90 mL H<sub>2</sub>O and fractions of 5 mL were collected and identified by spotting on TLC plates and visualization *via* UV lamp. The fractions containing the nucleotide were combined and evaporated under reduced pressure. The residue was dissolved in 10 mL H<sub>2</sub>O and the absorbance of an aliquot of a 1:10 dilution was measured at 260 nm. The concentration was calculated *via* Beer-Lambert law (section 6.1) and stored at -20°C until further usage the nucleotide solution was stored.

#### **General procedure II.**

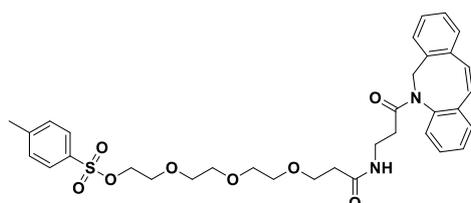
Synthesis of  $\gamma$ -O linked ATP analogues. The syntheses were performed in reference to Hacker and co-workers with adjustments.<sup>[98, 100]</sup> For each reaction, 100-300  $\mu$ mol of ATP Bu<sub>4</sub>N<sup>+</sup> salt were evaporated under reduced pressure, dissolved in 5 mL anhydrous DMF and evaporated under reduced pressure. The step of dissolving and subsequent evaporation was repeated two times and the final residue was dried on high vacuum overnight. 2.7 - 6.0 equiv. of alkylating agent were dissolved in 2-6 mL anhydrous DMF, then 1-2 g oven-dried 4 Å molecular sieves were added and stirred at room temperature overnight under Ar atmosphere. The next day the dried nucleotide was dissolved in 2-6 mL anhydrous DMF, flushed with Ar and equipped with a septum and Ar balloon to guarantee anhydrous conditions. The alkylating agent in DMF was then, under rigorous stirring, added dropwise, to the ATP solution and stirred at room temperature (r.t.) overnight. Reaction control was performed *via* analytical RP-HPLC (section 6.1). Unless no further conversion was monitored the solvent was evaporated under reduced

pressure. The residue was dissolved in 10 mL HPLC grade H<sub>2</sub>O and centrifuged (Hermle Labortechnik Centrifuge Z 206A) at 6000 rpm for 10 min at r.t. The supernatant was filtered and purified two to three times *via* semi-preparative RP-HPLC (section 6.1). The product fractions were identified by ESI-MS and lyophilized. The  $\gamma$ -O linked ATP tris-triethylammonium salts were, unless further specified, transferred into their sodium salt, either by precipitation or by ion exchange followed by dialysis. The conversion yield [%] of the isolated ATP analogues was determined by analytical RP-HPLC.

### General procedure III.

Conversion of  $\gamma$ -modified ATP tris-triethylammonium salts into their sodium salt form was performed by precipitation. Therefore, the lyophilized ATP tris-triethylammonium salt (ca. 2 mg) was dissolved in 200  $\mu$ L HPLC grade MeOH and added dropwise solution of anhydrous NaClO<sub>4</sub> in HPLC grade acetone (32 mg in 1.8 mL). After the mixture has been stored in the in the -20°C fridge for ca. 30 min colorless solid precipitated. The supernatant was removed and the pellet was washed three times with 1 mL HPLC grade acetone. The colorless solid was dried in high vacuum.

#### 6.3.1.1 Synthesis of 3-(2-(2-(2-(2-tosyloxyethoxy)ethoxy)ethoxy)ethoxy)-N-[3-oxo-N-(DBCO)propyl] propanamid (**24**)



**Chemical Formula:** C<sub>34</sub>H<sub>38</sub>N<sub>2</sub>O<sub>8</sub>S

**Molecular Weight:** 634.74

A solution of 3-(2-(2-(2-(2-tosyloxy)ethoxy)ethoxy)ethoxy)propanoic acid (108.41 mg, 288.0  $\mu$ mol, 1.0 equiv.) in anhydrous DMF (2.88 mL, 10 mL/mmol) and NMM (34.96 mg, 38.00  $\mu$ L, 345.6  $\mu$ mol, 1.2 equiv.) was stirred at -15°C for 5 min, before IBCF (43.27 mg, 41.09  $\mu$ L, 316.80  $\mu$ mol, 1.1 equiv.) was added. The solution was stirred for 15 min and then dibenzocyclooctyne-amine (87.54 mg, 0.316.80  $\mu$ mol; 1.1 equ.), dissolved in anhydrous DMF (4 mL), was added dropwise to the reaction mixture. After stirring overnight under Ar atmosphere and allowing to warm up to r.t., reaction control was performed *via* TLC. The

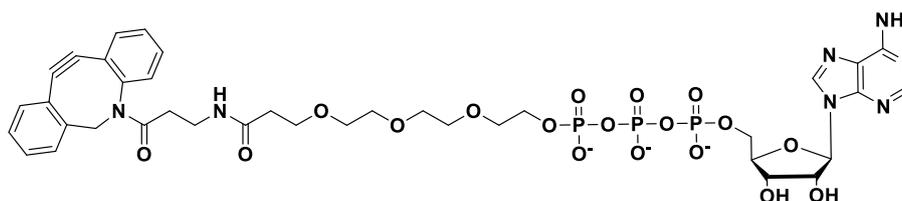
solvent was evaporated under reduced pressure and the residue was dissolved in EtOAc (50 mL) and washed two times with *sat.* NaHCO<sub>3</sub> solution and subsequently with brine (50 mL). The aqueous layers were extracted with EtOAc and all combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and the solvent was evaporated under reduced pressure. The crude product was purified *via* silica column chromatography (EtOAc:MeOH/20:1, v/v) obtaining **24** (151.72 mg, 239 μmol, 83%) as yellow oil.

**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>) δ 7.78 (d, 2H, Ar), 7.65 (d, *J* = 7.9 Hz, 1H, Ar), 7.41 – 7.28 (m, 8H, Ar), 6.56 (t, *J* = 6.2 Hz, 1H, Ar), 5.12 (d, *J* = 13.9 Hz, 1H, DBCO-CH<sub>2</sub>), 4.12 (t, *J* = 5.1 Hz, 2H, CH<sub>2</sub>-Tos), 3.72 – 3.43 (m, 13H, DBCO-CH<sub>2</sub>, 6xCH<sub>2</sub>), 3.36 – 3.20 (m, 2H, CH<sub>2</sub>-NH), 2.53 – 2.45 (m, 1H, N-CO-CH<sub>2</sub>-CH<sub>2</sub>-NH-CO), 2.43 (s, 3H, CH<sub>3</sub>), 2.36 – 2.25 (m, 2H, PEG-CH<sub>2</sub>-CO-NH), 1.98 – 1.91 (m, 1H, N-CO-CH<sub>2</sub>-CH<sub>2</sub>-NH-CO). **<sup>13</sup>C NMR** (126 MHz, CDCl<sub>3</sub>) δ 172.09, 171.14, 151.16, 148.16, 144.92, 133.00, 132.19, 129.94, 129.19, 128.70, 128.45, 128.35, 128.08, 127.93, 127.30, 125.68, 123.10, 122.58, 114.82, 107.88, 70.83, 70.55, 70.47, 70.30, 69.36, 68.75, 67.19, 55.57, 36.93, 35.23, 34.84, 32.02, 29.47, 21.76.

**HRMS** (ESI): *m/z* calcd. for C<sub>34</sub>H<sub>39</sub>N<sub>2</sub>O<sub>8</sub>S<sup>+</sup> [M+H]<sup>+</sup>: 635.2422, found: 635.2433

**R<sub>f</sub>** = 0.22 (SiO<sub>2</sub>; EtOAc/MeOH, 20:1).

### 6.3.1.2 Synthesis of γ-DBCO-PEG4-ATP (**25**)



**Chemical Formula:** C<sub>37</sub>H<sub>43</sub>N<sub>7</sub>O<sub>18</sub>P<sub>3</sub><sup>3-</sup>

**Molecular Weight:** 966.71

The synthesis of compound **25** was performed with ATP tetrabutylammonium salt (23.4 mg, 19 μmol, 1.0 equiv.), and **24** (36.18 mg, 57 μmol, 3.0 equiv.) as described in general procedure I and II, without the use of mol sieves. Semi preparative RP-HPLC purification afforded the triethylammonium salt of **25** as a colorless oil after lyophilization (31%). For subsequent

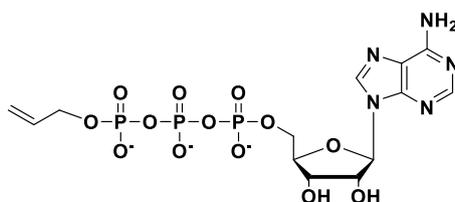
incorporation as pseudo substrate in the T4 PNK catalyzed reaction, **25** was transferred into its sodium salt by precipitation, as described in general procedure III.

**<sup>1</sup>H NMR** (500 MHz, MeOD)  $\delta$  8.61 (s, 1H, H2), 8.19 (s, 1H, H8), 7.66 (d,  $J$  = 6.7 Hz, 1H, Ar), 7.53 – 7.41 (m, 4H, Ar), 7.41 – 7.30 (m, 2H, Ar), 7.29 – 7.21 (m, 1H, Ar), 6.10 (d,  $J$  = 6.0 Hz, 1H, H1'), 5.14 (d,  $J$  = 14.1 Hz, 1H, DBCO-CH<sub>2</sub>), 4.73 (t,  $J$  = 5.5 Hz, 1H, H2'), 4.62 – 4.56 (m, 1H, H3'), 4.38 – 4.09 (m, 5H, H4', H5', CH<sub>2</sub>-O-P), 3.73 – 3.43 (m, 13H, DBCO-CH<sub>2</sub>, 6xCH<sub>2</sub>), 3.10 – 2.98 (m, 2H, CH<sub>2</sub>-NH), 2.56 – 2.44 (m, 1H, N-CO-CH<sub>2</sub>-CH<sub>2</sub>-NH-CO), 2.33 – 2.27 (m, 2H, PEG-CH<sub>2</sub>-CO-NH), 2.06 – 2.01 (m, 1H, N-CO-CH<sub>2</sub>-CH<sub>2</sub>-NH-CO). **<sup>13</sup>C NMR** (201 MHz, MeOD)  $\delta$  174.01, 173.21, 153.75, 152.61, 151.04, 149.49, 141.29, 133.47, 130.56, 130.07, 129.66, 129.22, 128.93, 128.11, 126.50, 124.37, 123.64, 120.11, 115.59, 108.85, 88.59, 76.27, 72.01, 71.93, 71.39, 71.31, 71.28, 71.19, 68.20, 57.58, 57.48, 57.37, 56.58, 43.09, 37.52, 36.69, 35.44. **<sup>31</sup>P NMR** (202 MHz, MeOD)  $\delta$  -11.10 (d,  $J$  = 19.1 Hz, 1P,  $\alpha$ ), -11.43 (d,  $J$  = 19.1 Hz, 1P,  $\gamma$ ), -22.86 (t,  $J$  = 19.1 Hz, 1P,  $\beta$ ).

**HRMS** (ESI):  $m/z$  calcd. for C<sub>37</sub>H<sub>45</sub>N<sub>7</sub>O<sub>18</sub>P<sub>3</sub><sup>-</sup> [M-H]<sup>-</sup>: 968.2039, found: 968.2044

**Analytical RP-HPLC** R<sub>t</sub> (gradient:0-60% buffer B) = 5.699 min

### 6.3.1.3 Synthesis of $\gamma$ -allyl-ATP (**26**)



**Chemical Formula:** C<sub>13</sub>H<sub>17</sub>N<sub>5</sub>O<sub>13</sub>P<sub>3</sub><sup>3-</sup>

**Molecular Weight:** 544.22

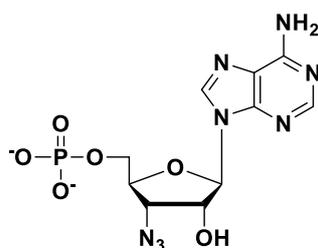
The synthesis of compound **26** was performed with ATP tetrabutylammonium salt (123.45 mg, 100  $\mu$ mol, 10 equiv.), and allyl bromide (65.48 mg, 541.28  $\mu$ mol, 5.4 equiv.) as described in general procedure I and II. RP-HPLC purification afforded the tris-triethylammonium salt of **26**, which was further transferred to its sodium salt (general procedure III) and isolated as a colorless solid after lyophilization (55%).

**<sup>1</sup>H NMR** (800 MHz, MeOD) δ 8.63 (s, 1H, H2), 8.17 (s, 1H, H8), 6.10 (d, *J* = 5.3 Hz, 1H, H1'), 5.97 (ddt, *J* = 17.2, 10.6, 5.3 Hz, 1H, CH-CH<sub>2</sub>), 5.32 – 5.27 (m, 1H, CH-CH<sub>2</sub>), 5.09 – 5.04 (m, 1H, CH-CH<sub>2</sub>), 4.66 (t, *J* = 5.1 Hz, 1H, H2'), 4.54 (dd, *J* = 5.0, 3.6 Hz, 1H, H3'), 4.51 – 4.46 (m, 2H, CH<sub>2</sub>-CH-CH<sub>2</sub>), 4.28 – 4.20 (m, 3H, H4', H5'). **<sup>13</sup>C NMR** (201 MHz, MeOD) δ 158.21, 154.61, 150.03, 143.04, 136.22, 120.72, 115.65, 87.29, 84.65, 76.49, 72.83, 67.88, 65.92. **<sup>31</sup>P NMR** (202 MHz, MeOD) δ -9.42 (d, *J* = 16.5 Hz, 1P, α), -9.88 (d, *J* = 16.8 Hz, 1P, γ), -19.87 (t, *J* = 16.6 Hz, 1P, β).

**HRMS** (ESI): *m/z* calcd. for C<sub>13</sub>H<sub>19</sub>N<sub>5</sub>O<sub>13</sub>P<sub>3</sub><sup>-</sup> [M-H]<sup>-</sup>: 546.0197, found: 546.0197

**Analytical RP-HPLC** R<sub>t</sub> (gradient:0-10% buffer B) = 4.420 min

#### 6.3.1.4 Synthesis of 3'-azido-2',3'-dideoxyadenosine monophosphate (**27**)



**Chemical Formula:** C<sub>10</sub>H<sub>11</sub>N<sub>8</sub>O<sub>6</sub>P<sup>2-</sup>

**Molecular Weight:** 370.06

The synthesis was performed following an adjusted protocol of Zhang and co-workers.<sup>[114], [207]</sup> Therefore, 3'-azido-2',3'-dideoxyadenosine (20 mg, 68.43 μmol, 1.0 equiv.), was added to a round bottom flask, which was previously heated under reduced pressure and floated with Ar. The flask was equipped with a septum and an Ar balloon and after triethyl phosphate (2.8 mL) was added *via* syringe, the mixture was stirred at ca. 1°C for 15 min, before POCl<sub>3</sub> (62.95 mg, 38 μL, 410.6 μmol, 6.0 equiv) was added dropwise *via* syringe. Since the reaction control, which was performed *via* analytical HPLC, indicated only 60% product formation, again POCl<sub>3</sub> (62.95 mg, 38 μL, 410.6 μmol, 6.0 equiv.) was added to the reaction mixture, while keeping the temperature at 1°C. After a total reaction time of 5.5 h, the reaction was quenched with 15 mL TEAA buffer (0.1 M). The reaction mixture was then washed with EtOAc (to remove P(OEt)<sub>3</sub>), and the combined aqueous solution was filtered. Semi preparative RP-HPLC purification afforded the tris-triethylammonium salt of **27** as a colorless oil after lyophilization (conversion yield determined *via* analytical RP-HPLC; 96%). **<sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O)** δ 8.51 (s,

1H, H2), 8.24 (s, 1H, H8), 6.11 (d, J = 5.8 Hz, 1H, H1'), 5.04 (t, J = 5.7 Hz, 1H, H2'), 4.53 (dd, J = 5.6, 4.0 Hz, 1H, H3'), 4.43 – 4.37 (m, 1H, H4'), 4.15 – 4.06 (m, 2H, H5'). <sup>31</sup>P NMR (202 MHz, D2O) δ -0.00.

**HRMS (ESI):** *m/z* calcd. for C<sub>10</sub>H<sub>12</sub>N<sub>8</sub>O<sub>6</sub>P<sup>-</sup> [M-H]<sup>-</sup>: 371.0623, found: 371.06168

**Analytical RP-HPLC R<sub>t</sub>** (gradient:0-10% buffer B) = 4.051 min

### 6.3.2 T4 PNK Catalyzed 5'-End Labelling Reaction

For analyzing compound **25** and **26** regarding their properties as potential T4 PNK substrates the reactions were performed in reference to Schönegger *et al.*<sup>[87]</sup> on a 43 mer oligonucleotide ODN1 (5'-3': GCC ATA GCG ATA GCG ATA CGC ACA CCC GGT CCT CCT AAT TTA T ) as follows: ODN1 (5 μL, 10 μM dilution in RNase free H<sub>2</sub>O, 50 pmol, 1.0 equiv.) was heat denatured at 70°C for 5 min and then immediately put on ice. After cooling to r.t. for 5 min, PEG6000 (10.4 μL of 24% solution, purchased from Thermo Scientific™), T4 PNK buffer (10X reaction buffer A, purchased from Thermo Scientific™), compound **25** and **26** (5 μL, 10 mM in H<sub>2</sub>O, 50 nmol, 1000 equiv.) respectively and T4 PNK (5 μL, 10U/μL, purchased from Thermo Scientific™), were added. The reaction mixtures (50 μL total volume) were mixed by pipetting and incubated at 37°C for 3 h. The samples were purified by spin column purification (Oligo Clean & Concentrator Kit from Zymo Research), the concentrations of the purified samples were determined on a Qubit 4 Fluorometer using the Qubit™ ssDNA Assay-Kit and subsequent analytical RP-HPLC analysis were performed as described in section 6.1. The results and relevant RP-HPL chromatograms are displayed in Table 1.

**Table 1.** RP-HPL chromatograms of ODN1 unmodified and purified reaction mixtures after T4 PNK catalyzed labelling reactions with compound **25** and **26** respectively. ODN1:

sample		analytical RP-HPL chromatogram (gradient: 0-30% buffer B)
ODN1		<p>RT<sub>ODN1</sub> = 4.523 min</p>
compound	conversion yield of 5'-end labelling of ODN1 with $\gamma$ -phosphate modified ATPs [%]	analytical RP-HPL chromatogram (gradient: 0-30% buffer B)
<b>25</b>	0	<p>RT<sub>ODN1</sub> = 4.558 min</p>
<b>26</b>	0	<p>RT<sub>ODN1</sub> = 4.597 min</p>

## 6.4 Click-Modified Cu(II)-Dependent Nucleic Acid-Based Artificial Ribonuclease for Targeted Cleavage of SARS-CoV-2 Genomic RNA

### 6.4.1 Synthesis of Internal C8-Alkyne Modified Oligonucleotides

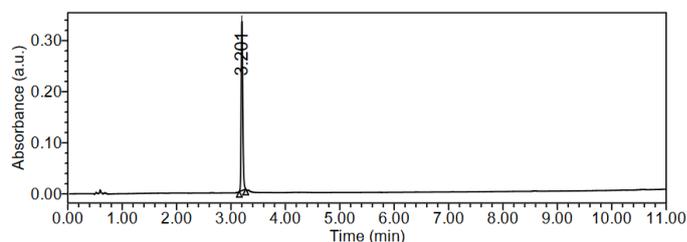
Internal C8-alkyne modified oligonucleotides were either synthesized on 1  $\mu\text{mol}$  scale using DNA SynBase™ CPG 1000/110 as solid supports on an automated DNA synthesizer (K&A H-8 SE Oligo-Synthesizer) with standard phosphoramidite chemistry, or otherwise provided by Ella Biotech GmbH using proprietary methods. The phosphoramidites of canonical deoxyribonucleosides (Bz-dA, Bz-dC, iBu-dG, dT) were purchased from Glen Research or Sigma-Aldrich, C8-alkyne-dT-CEP was provided by baseclick GmbH and solid supports were purchased from Link Technologies Ltd (DNA SynBase™ CPG 1000/110). The coupling efficiency was monitored with a trityl monitor. For the synthesis, DCA in  $\text{CH}_2\text{Cl}_2$  was employed as deblocking agent, Activator42® as an activator,  $\text{Ac}_2\text{O}$  in pyridine/THF as capping reagent and  $\text{I}_2$  as oxidizer in pyridine/ $\text{H}_2\text{O}$ . The cleavage and deprotection of the CPG bound oligonucleotides and deprotection was carried out by treating the solid support with 30%  $\text{NH}_4\text{OH}$  (1 mL) at r.t. overnight (300 rpm on the thermomixer). After centrifugation the supernatant was collected, and the beads were washed three times with 0.5 mL  $\text{H}_2\text{O}$ . The combined aqueous solutions were concentrated under reduced pressure for 1.5 h at 30°C using a SpeedVac concentrator (Savant™ SpeedVac™ SPD120 Vacuum Concentrator) and the remaining residue was lyophilized. The oligonucleotides were further purified by semi-preparative RP-HPLC (section 6.1) using a gradient of 0-40% of buffer B over 45 min was applied for purification. The purified oligonucleotides were analyzed by analytical RP-HPLC (section 6.1) using gradient was 0-30% buffer B from 0  $\rightarrow$  4 min, followed by 30-85% buffer B from 4  $\rightarrow$  10 min. The concentrations of the oligonucleotides were calculated as described in section 6.1 and the structural integrity of the synthesized oligonucleotides was determined by MALDI-TOF mass measurements (section 6.1, Table 2). The isolated yield of the oligonucleotides after HPLC purification was between 290 and 360 nmol (29-36%).

**Table 2.** left) RP-HPL-chromatograms (gradient: 0-30% buffer B from 0 → 4 min, followed by 30-85% buffer B from 4 → 10 min) and right) MALDI-TOF mass spectra (positive mode) of purified oligonucleotides.

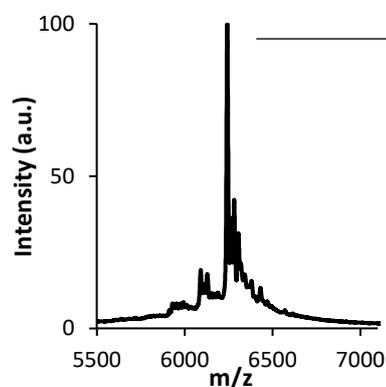
Sequence: **ODN2**

5'- ACT AAG CAX GCA GCC GAG TG -3'

R<sub>t</sub>= 3.201 min



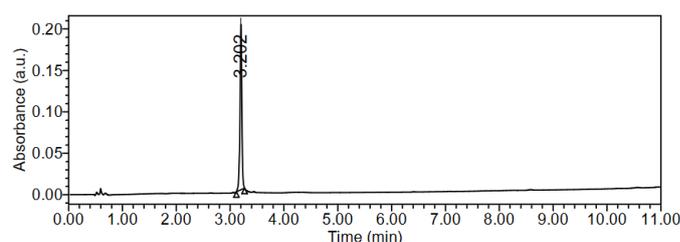
[M+H]<sup>+</sup> m/z calcd. 6242.18  
found 6242.64



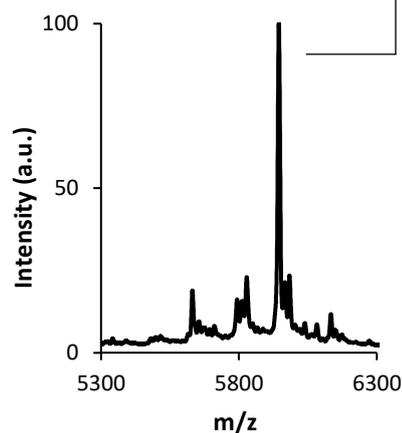
Sequence: **ODN3A**

5'- TGC GTG AGX GCA CTA AGC A -3'

R<sub>t</sub>= 3.202 min



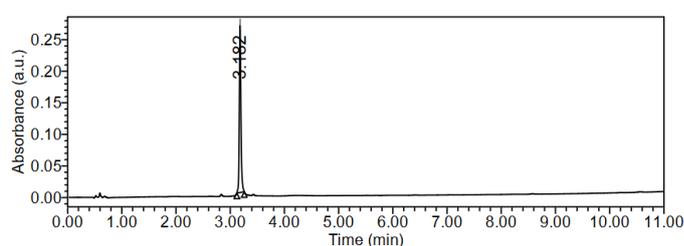
[M+H]<sup>+</sup> m/z calcd. 5943.98  
found 5944.80



Sequence: **ODN3C**

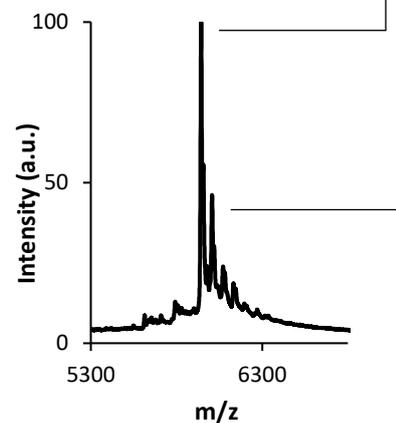
5'- TGC GXG AGT GCA CTA AGC A -3'

R<sub>t</sub>= 3.182 min



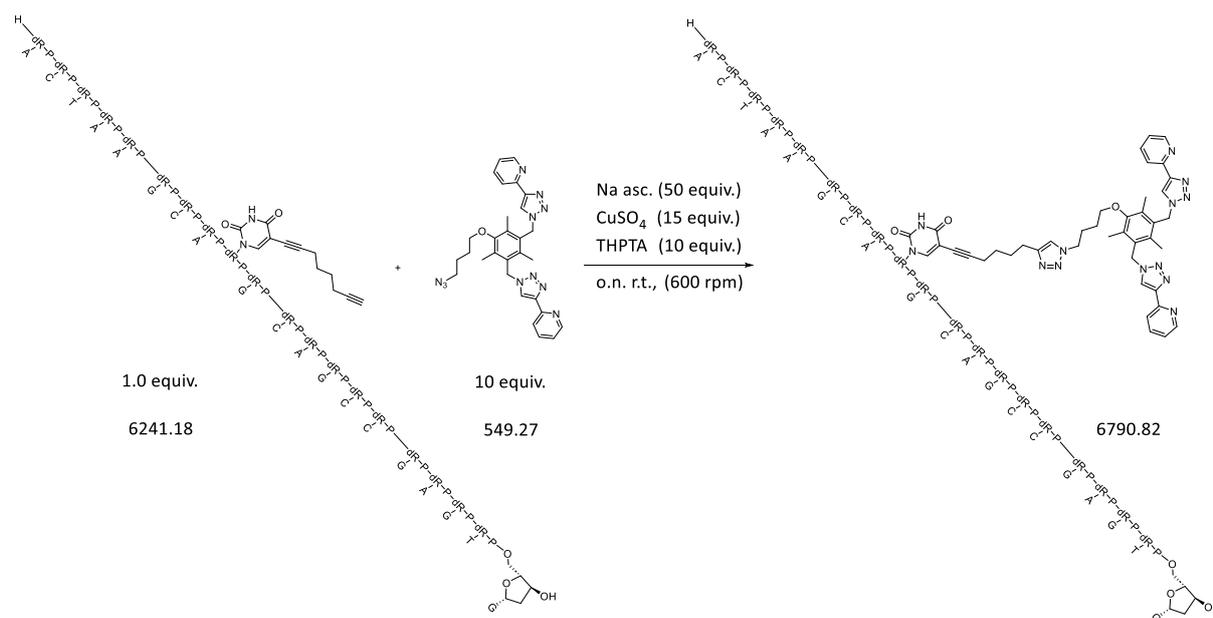
[M+H]<sup>+</sup> m/z calcd. 5943.98  
found 5943.68

[M-H+K+Na]<sup>+</sup> calcd. 6004.06  
found 6005.88



## 6.4.2 CuAAC of N<sub>3</sub>-Cu-Ligand to Alkyne Modified Oligonucleotides

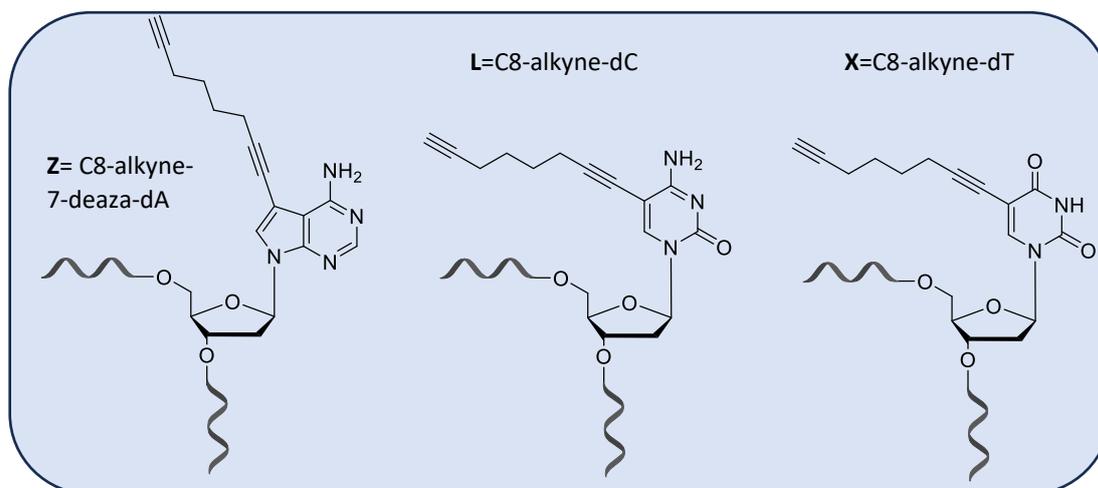
The copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) of the C8-alkyne modified oligonucleotides (Table 3, Figure 45 ) and the DC-Pyr-N<sub>3</sub> ligand (chemical formula: C<sub>29</sub>H<sub>31</sub>N<sub>11</sub>O) were performed following an adjusted protocol described by Lauria *et al.*<sup>[151]</sup> The CuAAC reactions were performed in two different scales (25 nmol and 250 nmol). The CuAAC click reaction is displayed in Scheme 8, on the example of ODN2.



**Scheme 8.** Schematic representation of the CuAAC on ODN2 with DC-Py-N<sub>3</sub> ligand (exemplary scheme represents all CuAAC reactions). dR= deoxyribose backbone, P= phosphate, A, C, G, T= nucleobases. Oligonucleotide sequences are displayed using the HELM Monomer Toolbar from Chemdraw 22.0.0.

**Table 3.** Internal C8-alkyne modified oligonucleotides.

Name	Sequence (5' to 3')	Modification	Length (nt)	$\epsilon$ [L/(mole·cm)]
ODN1A	AGC CGA GXG ACA GCC ACA CA	X=C8-Alkyne-dT	20	198900
ODN1B	AGC CGA GTG ALA GCC ACA CA	L=C8-Alkyne-dC	20	201400
ODN1C	AGC CGA GTG ACA GCC ALA CA	L=C8-Alkyne-dC	20	201400
ODN2	ACT AAG CAX GCA GCC GAG TG	X=C8-Alkyne-dT	20	197100
ODN3A	TGC GTG AGX GCA CTA AGC A	X=C8-Alkyne-dT	19	185000
ODN3B	TGC GTG AGT GCA CTZ AGC A	Z=C8-Alkyne-7-deaza-dA	19	186300
ODN3C	TGC GXG AGT GCA CTA AGC A	X=C8-Alkyne-dT	19	185000
ODN4	AGT GCA CTA AGC AXG CAG CCG AGT GAC AGC C	X=C8-Alkyne-dT	31	302900
ODN5	GCA TGC AGC CGA GXG ACA GC	X=C8-Alkyne-dT	20	193600
ODN6	GAG TGC ACX AAG CAT GCA GC	X=C8-Alkyne-dT	20	196600

**Figure 45.** Schematic representation of internal C8-alkyne modifications.

**CuAAC Small Scale (25 nmol).** In a 1.5 mL Eppendorf tube, 250  $\mu$ l of alkyne modified oligonucleotide (100  $\mu$ M in H<sub>2</sub>O, 25 nmol, 1 equiv.), was added to 25  $\mu$ l of DC-Py-N<sub>3</sub> ligand (10 mM in DMSO, 250 nmol, 10 equiv.) and degassed for 10 min with Ar. Then 225  $\mu$ l of DMSO (degassed for 10 min with Ar) was added to get a 1:1 H<sub>2</sub>O: DMSO dilution. In separate 1.5 mL Eppendorf tube the catalytic solution was prepared as follows: 25  $\mu$ l of THPTA (10 mM in H<sub>2</sub>O, 250 nmol, 10 equiv.), 37.5  $\mu$ l of CuSO<sub>4</sub> (10 mM in H<sub>2</sub>O, 375 nmol, 15 equiv.) 12.5  $\mu$ l of Na asc. (100 mM, 1250 nmol, 50 equiv.). The catalytic solution was degassed for 10 min with Argon,

then the combined with oligo/DC-Py-N<sub>3</sub> ligand solution and mixed on the thermocycler at r.t. with 600 rpm overnight. After reaction control *via* HPLC, the reaction was quenched with 225  $\mu$ L EDTA (50 mM in H<sub>2</sub>O, 11250 nmol, 450 equiv.). The crude reaction mixture was desalted using Glen Gel-Pak™ 1.0 Desalting Column according to the manufacturer's procedure. For subsequent RP-HPLC purification the desalted reaction mixture was concentrated to a volume of 100  $\mu$ L using a SpeedVac concentrator. The clicked oligonucleotides were further purified by semi preparative RP-HPLC (section 6.1). A gradient of 0-40% of buffer B over 45 min was applied for purification. The collected fractions were lyophilized, rediluted in 1 mL H<sub>2</sub>O and again desalted using Glen Gel-Pak™ 1.0 Desalting Column according to the manufacturer's procedure. The purified oligonucleotides were analyzed by analytical RP-HPLC (section 6.1). A gradient of 0-30% buffer B from 0  $\rightarrow$  4 min, 30-85% buffer B from 4  $\rightarrow$  10 min was applied. The concentrations of the oligonucleotides were calculated as described in section 6.1, whereas for the extinction coefficient ( $\epsilon$ ) of the Cu-ligand clicked oligonucleotide, the same value as for the corresponding alkyne modified oligonucleotide was chosen, since the impact of the clicked DC-Py-N<sub>3</sub> ligand is supposed to be negligible. The structural integrity of the ligand-clicked oligonucleotides (ODN1-6\*) were determined by MALDI-TOF mass measurements (section 6.1, Table 4).

**CuAAC Large Scale (250 nmol).** In a 10 mL round flask equipped with a magnetic stirring bar, 250  $\mu$ L DC-Py-N<sub>3</sub> ligand (10 mM in DMSO, 10 equiv., 2500 nmol), 250  $\mu$ L of alkyne modified DNA oligonucleotide (100  $\mu$ M in H<sub>2</sub>O, 250 nmol, 1 equiv.) and 2995  $\mu$ L DMSO (to reach a final 1:1, DMSO: H<sub>2</sub>O dilution) were combined and degassed with Ar for 10 min. In a falcon the catalytic solution was prepared as follows: 250  $\mu$ L of THPTA (10 mM in H<sub>2</sub>O, 2500 nmol, 10 equiv.), 375  $\mu$ L of CuSO<sub>4</sub> (10 mM in H<sub>2</sub>O, 3750 nmol, 15 equiv.), 125  $\mu$ L of Na asc. (100 mM, 12500 nmol, 50 equiv.). The catalytic solution was degassed for 10 min with Ar, then the combined with oligo-DC-Py-N<sub>3</sub> ligand solution and stirred at r.t. overnight. After reaction control *via* HPLC, the reaction was quenched with 1125  $\mu$ L EDTA (100 mM in H<sub>2</sub>O, 11250 nmol, 450 equiv.). The crude reaction mixture was concentrated to 1 mL using a SpeedVac concentrator and then desalted using Glen Gel-Pak™ 1.0 Desalting Column according to the manufacturer's procedure. The desalted clicked oligonucleotides (3.5-4 mL) were purified by semi-preparative RP-HPLC. Further work up and characterization steps were performed as described in the CuAAC-Small scale protocol. The conversion yield [%] of the Cu-binding ligand

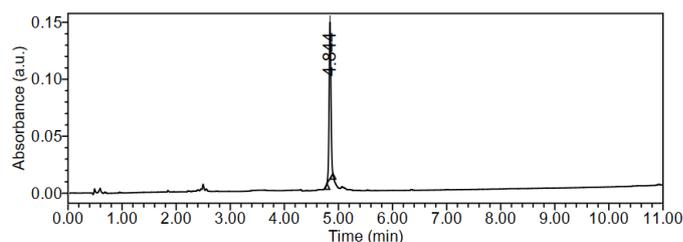
clicked oligonucleotides (ODN1-6\*) was determined by analytical RP-HPLC, varying between 95 and 99 %.

**Table 4.** left) RP-HPL-chromatograms (gradient: 0-30% buffer B from 0 → 4 min, followed by 30-85% buffer B from 4 → 10 min) and right) MALDI-TOF mass spectra (positive mode) of purified ligand clicked oligonucleotides.

Sequence: **ODN1A\***

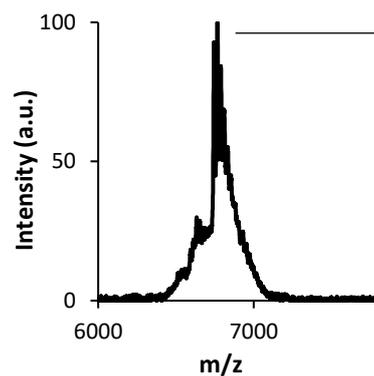
5'- AGC CGA GX\*G ACA GCC ACA CA -3'

R<sub>t</sub>= 4.844 min



[M+H]<sup>+</sup> m/z calcd. 6745.81

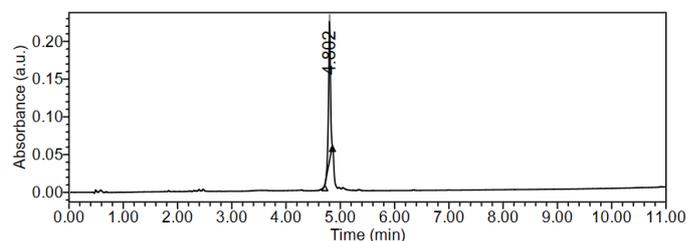
found 6744.932



Sequence: **ODN1B\***

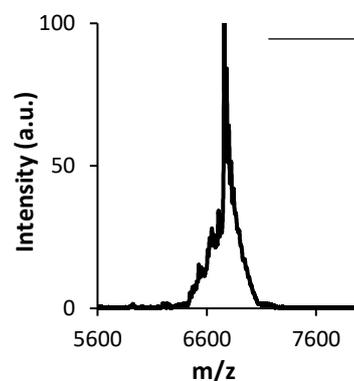
5'- AGC CGA GTG AL\*A GCC ACA CA -3'

R<sub>t</sub>= 4.802 min



[M+H]<sup>+</sup> m/z calcd. 6759.83

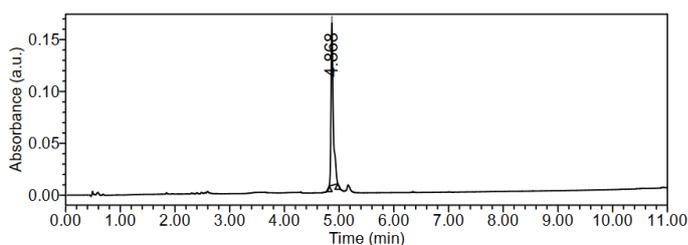
found 6759.98



Sequence: **ODN1C\***

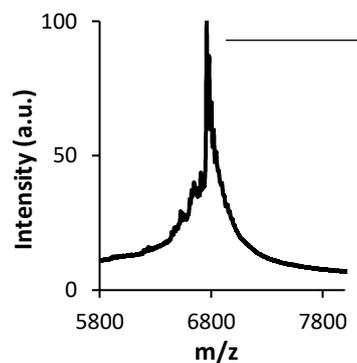
5'- AGC CGA GTG ACA GCC AL\*A CA -3'

R<sub>t</sub>= 4.868 min



[M+H]<sup>+</sup> m/z calcd. 6759.83

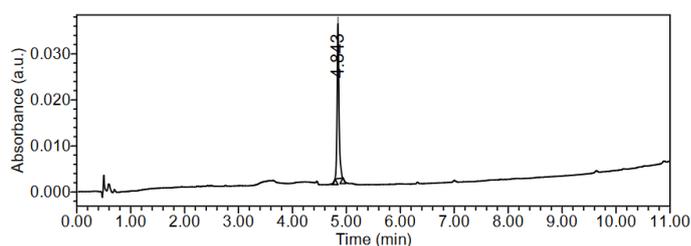
found 6759.58



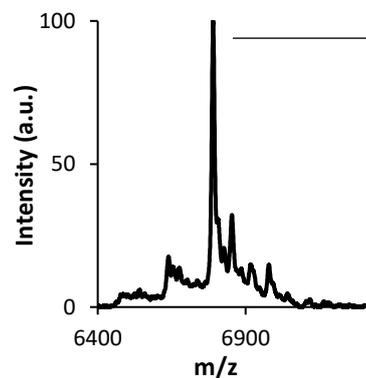
Sequence: **ODN2\***

5'- ACT AAG CAX\* GCA GCC GAG TG -3'

R<sub>t</sub>= 4.843 min



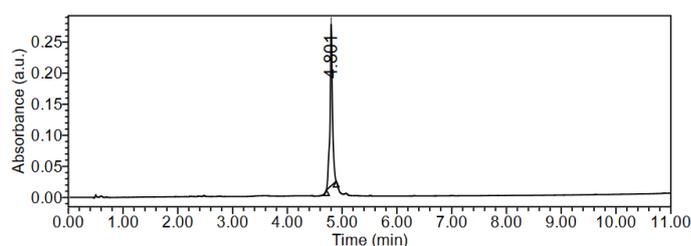
[M+H]<sup>+</sup> m/z calcd. 6790.82  
found 6790.12



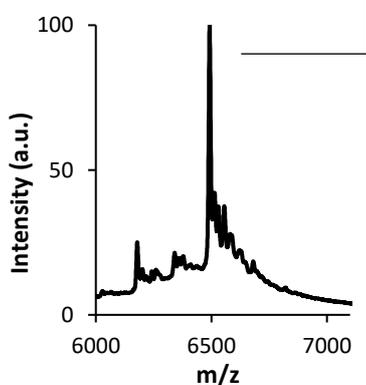
Sequence: **ODN3A\***

5'- TGC GTG AGX\* GCA CTA AGC A -3'

R<sub>t</sub>= 4.801 min



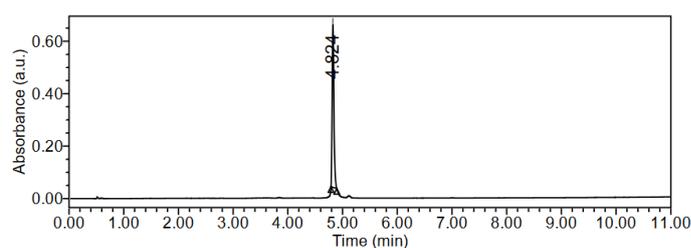
[M+H]<sup>+</sup> m/z calcd. 6493.63  
found 6492.52



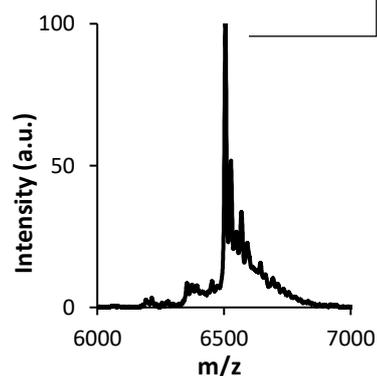
Sequence: **ODN3B\***

5'- TGC GTG AGT GCA CTZ\* AGC A -3'

R<sub>t</sub>= 4.824 min



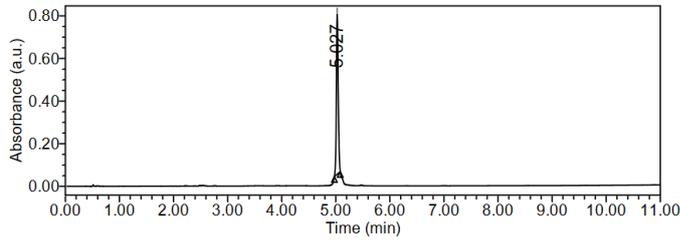
[M+H]<sup>+</sup> m/z calcd. 6506.67  
found 6505.34



Sequence: **ODN3C\***

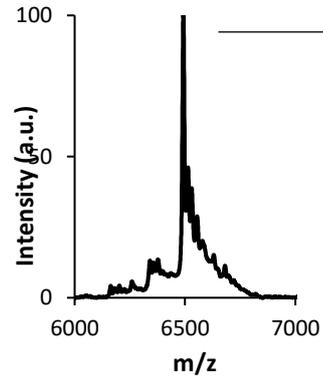
5'- TGC **GX\***G AGT GCA CTA AGC A -3'

R<sub>t</sub>= 5.027 min



[M+H]<sup>+</sup> m/z calcd. 6493.63

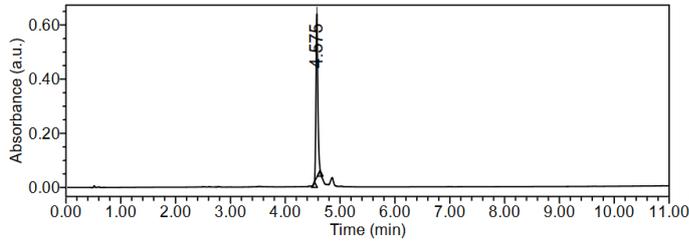
found 6492.52



Sequence: **ODN4\***

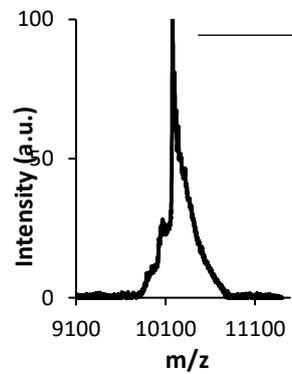
5'- AGT GCA CTA AGC **AX\***G CAG CCG AGT GAC AGC C -3'

R<sub>t</sub>= 4.575 min



[M+H]<sup>+</sup> m/z calcd. 10180.01

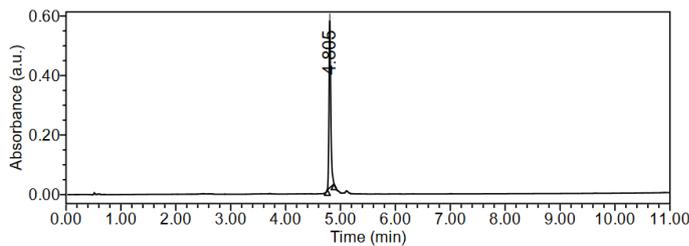
found 10178.93



Sequence: **ODN5\***

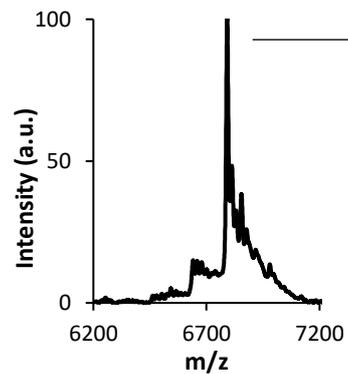
5'- GCA TGC AGC CGA **GX\***G ACA GC -3'

R<sub>t</sub>= 4.805 min

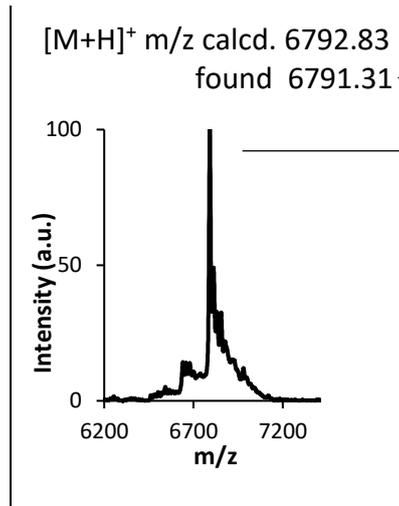
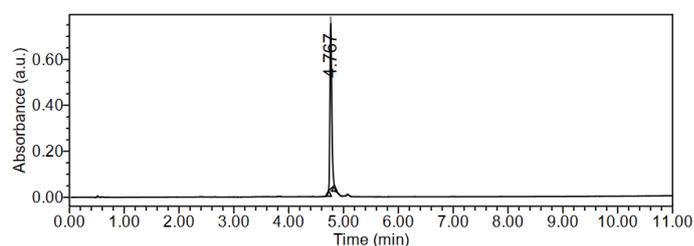


[M+H]<sup>+</sup> m/z calcd. 6792.81

found 6791.70



Sequence: **ODN6\***  
 5'- GAG TGC ACX\* AAG CAT GCA GC -3'  
 R<sub>t</sub>= 4.767 min



**Table 5.** Internal C8-alkyne modified and ligand-clicked (\*) oligonucleotides

Name	Sequence (5' to 3')	Modification	Length (nt)	$\epsilon$ [L/(mole·cm)]
ODN1A*	AGC CGA GX*G ACA GCC ACA CA	X=C8-Alkyne-dT	20	198900
ODN1B*	AGC CGA GTG AL*A GCC ACA CA	L=C8-Alkyne-dC	20	201400
ODN1C*	AGC CGA GTG ACA GCC AL*A CA	L=C8-Alkyne-dC	20	201400
ODN2*	ACT AAG CAX* GCA GCC GAG TG	X=C8-Alkyne-dT	20	197100
ODN3A*	TGC GTG AGX* GCA CTA AGC A	X=C8-Alkyne-dT	19	185000
ODN3B*	TGC GTG AGT GCA CTZ* AGC A	Z=C8-Alkyne-7-deaza-dA	19	186300
ODN3C*	TGC GX*G AGT GCA CTA AGC A	X=C8-Alkyne-dT	19	185000
ODN4*	AGT GCA CTA AGC AX*G CAG CCG AGT GAC AGC C	X=C8-Alkyne-dT	31	302900
ODN5*	GCA TGC AGC CGA GX*G ACA GC	X=C8-Alkyne-dT	20	193600
ODN6*	GAG TGC ACX* AAG CAT GCA GC	X=C8-Alkyne-dT	20	196600

\*=ligand-clicked

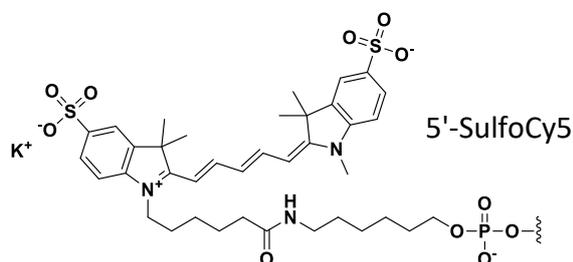
### 6.4.3 Binding Affinity Analysis of C8-Alkyne and Cu-Binding Ligand Clicked Oligonucleotides towards Target, Off-Target Sequences *via* nPAGE

Target and off target oligoribonucleotides were purchased from Ella Biotech GmbH and the sequences are displayed in Table 6.

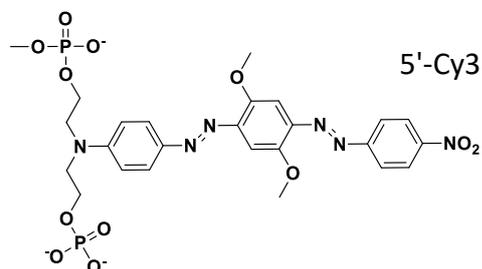
**Table 6.** Oligoribonucleotide target (T) and off target (OT) sequences

T/OF	Name	Type	Sequence (5' to 3')	Modification	Length (nt)
T	ORN1	RNA	CUG UGU GGC UGU CAC UCG GCU GCA UGC UUA GUG CAC UCA CGC AG	-	44
T	ORN2	modif. RNA	Z CUG UGU GGC UGU CAC UCG GCU GCA UGC UUA GUG CAC UCA CGC AG	5'-SulfoCy5=Z	44
T	ORN3	modif. RNA	Z CUG UGU GGC UGU CAC UCG GCU GCA UGC UUA GUG CAC UCA CGC AG-BHQ2-PHO	5'-SulfoCy5=Z 3'-BHQ2-PHO	44
OT	ORN4	RNA	CAC CGA GGC CAC GCG GAG UAC GAU CGA GUG UAC AGU G	-	37
OT	ORN5	modif. RNA	CY3-CAC CGA GGC CAC GCG GAG UAC GAU CGA GUG UAC AGU G-3'	5'-Cy3	37
OT	ORN6	modif. RNA	CY3-CAC CGA GGC CAC GCG GAG UAC GAU CGA GUG UAC AGU G-BHQ2-PHO	5'-Cy3 3'-BHQ2-PHO	37

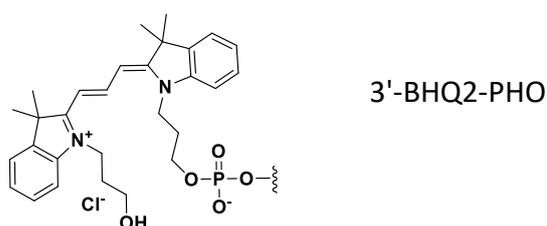
Structure of the Z modification is given below



Structure of the Cy3 modification is given



Structure of the BHQ2 (black hole quencher 2)-PHO modification is given below



#### 6.4.3.1 Binding Affinity Analysis of C8-Alkyne and Cu- Ligand Clicked Oligonucleotides towards Target, Off-Target without Cu(II) -Source

In a 200  $\mu\text{L}$  PCR vial C8-alkyne modified oligonucleotides ODN1-6 (0.5  $\mu\text{L}$  of 5  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 2.5 pmol, 5 equiv.) and ligand-clicked oligonucleotides ODN1-6\* (0.5  $\mu\text{L}$  of 5  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 2.5 pmol, 5 equiv.) respectively, were mixed with target ORN2 (0.5  $\mu\text{L}$  of 1  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 0.5 pmol, 1 equiv.), off-target ORN5 (0.5  $\mu\text{L}$  of 1  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 0.5 pmol, 1 equiv.) and 4.5  $\mu\text{L}$  RNA buffer (10 mM HEPES, 150 mM NaCl, 2 mM  $\text{MgCl}_2$ , pH 7). For the control target and off-target sample ORN2 and ORN5 respectively (0.5  $\mu\text{L}$  of 1  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 1 equiv.) were mixed with 4.5  $\mu\text{L}$  RNA buffer (10 mM HEPES, 150 mM NaCl, 2 mM  $\text{MgCl}_2$ , pH 7) and 1  $\mu\text{L}$  nuclease free  $\text{H}_2\text{O}$  in a 200  $\mu\text{L}$  PCR vial. All samples (6  $\mu\text{L}$  each) were incubated at 37°C in the thermocycler for 2 h and subsequently 1  $\mu\text{L}$  of Orange dye (6) was added to the samples and loaded onto the gel. Native PAGE (nPAGE, 20%, section 6.1 B) was run for 4 h at 70 V and afterwards imaged using Cy3 and Cy5 filters, before being stained for 10 min with SYBR Gold (1X dilution) and again imaged again using Cy3, Cy5 and blue filters (for visualization of SYBR gold stain).

#### 6.4.3.2 Binding Affinity Analysis of Selected Cu-Ligand-Clicked Oligonucleotides towards Target, Off-Target with Cu(II) Source

In a 200  $\mu\text{L}$  PCR vial ligand-clicked oligonucleotides ODN1A-C\*, ODN2\*, ODN4\* and ODN5\* (0.5  $\mu\text{L}$  of 5  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 2.5 pmol, 5 equiv.) respectively, were mixed with  $\text{Cu}(\text{Cl}_4\text{O})_2$  (0.5  $\mu\text{L}$  of 10  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 5.0 pmol, 10 equiv.) and incubated at r.t. for 20 min. Next, ORN2 (0.5  $\mu\text{L}$  of 1  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 0.5 pmol, 1 equiv.), ORN5 (0.5  $\mu\text{L}$  of 1  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 0.5 pmol, 1 equiv.) and 4.0  $\mu\text{L}$  RNA buffer (10 mM HEPES, 150 mM NaCl, 2 mM  $\text{MgCl}_2$ , pH 7) were added, mixed and incubated for 18 h at 37°C. For the control samples, ODN1A-C\*, ODN2\*, ODN4\* and ODN5\* (0.5  $\mu\text{L}$  of 5  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 2.5 pmol, 5 equiv.) were mixed with 4.5  $\mu\text{L}$  RNA buffer (10 mM HEPES, 150 mM NaCl, 2 mM  $\text{MgCl}_2$ , pH 7) and the control target and off-target samples were prepared as described in section 6.4.3.1. All control samples were incubated for 18 h at 37°C. Afterwards, nPAGE and imaging was performed as described in section 6.4.3.1.

#### 6.4.4 Cleavage Analysis of Cu-Ligand-Clicked Oligonucleotides towards Target, Off-Target

**Initial cleavage analysis.** ODN5\* (0.5 or 1  $\mu\text{L}$  of 5  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 2.5 or 5 pmol, 5 or 10 equiv.) was mixed with different amounts of  $\text{Cu}(\text{Cl}_4\text{O})_2$  (0.5 or 1  $\mu\text{L}$  of 10  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 5 or 10 pmol, 10 or 20 equiv.) and incubated at r.t. for 20 min. Then Na asc. (different amounts, 10-100  $\mu\text{M}$  in  $\text{H}_2\text{O}$ , 12.5-250 pmol, 25-500 equiv.) was added to the ODN- $\text{Cu}(\text{Cl}_4\text{O})_2$  -solution and incubated again at r.t. for 10 min. Afterwards ORN2 (0.5  $\mu\text{L}$  of 1  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 0.5 pmol, 1 equiv.), ORN5 (0.5  $\mu\text{L}$  of 1  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 0.5 pmol, 1 equiv.) and 0.5-2.75  $\mu\text{L}$  RNA buffer (10 mM HEPES, 150 mM NaCl, 2 mM  $\text{MgCl}_2$ , pH 7), to reach a total volume of 6  $\mu\text{L}$ , were added, mixed and incubated for 18 h at 37°C. The target and off-target control samples were prepared as described in 6.4.3.1 and also incubated for 18 h at 37°C. Afterwards, nPAGE and imaging were performed as described in section 6.4.3.1. The ratios of the different conditions can be taken from Table 7.

**Table 7.** Ratios of different conditions for the initial cleavage experiments. Numbers represent equivalents (1= 1 equiv. = 0.5 pmol).

conditions	compound				
	ORN2 (T)	ORN5 (OT)	ODN5*	$\text{Cu}(\text{Cl}_4\text{O})_2$	Na Asc.
A1	1	1	5	10	250
A2	1	1	5	10	25
B1	1	1	10	20	500
B2	1	1	10	20	50

**Cleavage analysis with increased Na ascorbate. concentration.** ODN5\* and ODN4\* (0.5 or 1  $\mu\text{L}$  of 5  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 2.5 or 5 pmol, 5 or 10 equiv.) respectively were mixed with different amounts of  $\text{Cu}(\text{Cl}_4\text{O})_2$  (0.5 or 1  $\mu\text{L}$  of 10  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 5 or 10 pmol, 10 or 20 equiv.) and incubated at r.t. for 20 min. Then Na asc. (different amounts, 1-10 mM in  $\text{H}_2\text{O}$ , 1250-5000 pmol, 2500-10000 equiv.) was added to the ODN- $\text{Cu}(\text{Cl}_4\text{O})_2$  -solution and incubated again at r.t. for 10 min. Afterwards, ORN2 (0.5  $\mu\text{L}$  of 1  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 0.5 pmol, 1 equiv.), ORN5 (0.5  $\mu\text{L}$  of 1  $\mu\text{M}$  dilution in  $\text{H}_2\text{O}$ , 0.5 pmol, 1 equiv.) and 1.5- 2.75  $\mu\text{L}$  RNA buffer (10 mM HEPES, 150 mM NaCl, 2 mM  $\text{MgCl}_2$ , pH 7), to reach a total volume of 6  $\mu\text{L}$ , were added, mixed and incubated for 18 h at 37°C. The target and off-target control samples were

prepared as described in section 6.4.3.1 and also incubated for 18 h at 37°C. Afterwards nPAGE and imaging were performed as described in section 6.4.3.1. The ratios of the different conditions can be taken from Table 8.

**Table 8.** Ratios of different conditions for the cleavage experiments with increased Na asc. Numbers represent equivalents (1= 1 equiv.= 0.5 pmol)

conditions	compound				
	ORN2 (T)	ORN5 (OT)	ODN5*	Cu(Cl <sub>4</sub> O) <sub>2</sub>	Na Asc.
A	1	1	5	10	<b>2500</b>
B	1	1	5	10	<b>5000</b>
C	1	1	10	20	<b>5000</b>
D	1	1	10	20	<b>10000</b>

#### 6.4.5 Binding Affinity Measurement and Determination of Dissociation Constant $K_D$ via Microscale Thermophoresis (MST)

Microscale thermophoresis (MST) was performed on a Monolith NT115 device using standard capillaries from NanoTemper Technologies GmbH. All device relevant consumables were purchased from NanoTemper Technologies GmbH and the instrument was calibrated according to the manufacturer's instructions. General settings for all MST experiments were applied as follows: thermostat setpoint: 25°C, excitation type- Nano-RED, excitation-power: 40%, MST-power: medium. The instructions for the sample preparation of the individual experiments were provided by the Monolith Control (MO) software. MST measurements were analyzed using the MO Analysis software (v3.0.5), and the generated data was exported and processed in Excel. The  $K_D$  values are generated by the MO Analysis'  $K_D$  model.

MST experiments were performed for C8-alkyne modified oligonucleotides, ODN2, ODN4, ODN5, and ligand-clicked oligonucleotides ODN2\*, ODN4\*, ODN5\*. All dilutions were prepared in RNA buffer (10 mM HEPES, 150 mM NaCl, 2 mM MgCl<sub>2</sub>, pH 7). Two different experimental series were performed, with a constant target (ORN2) concentration of 20 nM and varying ligand (ODN2, ODN4, ODN5, ODN2\*, ODN4\*, ODN5\*) concentrations with the highest concentration of either 400 nM or 2.5 μM. For the sample preparation a 1 μM stock solution of ORN2 in H<sub>2</sub>O was prepared and diluted in RNA buffer according to the MO control software. For the individual ligands (oligonucleotides) 800 nM or 5 μM stock solutions in

nuclease free H<sub>2</sub>O were prepared and again diluted in RNA buffer according to the MO control software. The preparation of a serial dilution for each experiment (16 varying ligand concentration and constant target concentration) was prepared following the MO control software, starting with the highest concentration of either 400 nM or 2.5 μM. Subsequently, to each sample of the serial dilution (16 dilutions in total), 10 μL of 40 nM target was added to reach a total volume of 20 μL, mixed and incubated for 4 h at 37°C. The samples were transferred to the MO capillaries and the measurement was started.

## 6.5 Investigations towards Copper-Free Click Chemistry for Optimized mRNA Vaccine Development

**General remarks.** ORN1 samples were purified by spin column purification (Oligo Clean & Concentrator Kit from Zymo Research), the concentrations of the purified samples were determined on a Qubit 4 Fluorometer using the Qubit™ RNA HS Assay-Kit.

### 6.5.1 Quantification of 3'-N<sub>3</sub> Labelling and SPAAC with DBCO-Trimannose-Moiety for baseclick's mRNA vaccine Development

#### 3'-N<sub>3</sub> labelling of ribo nucleic acid and SPAAC with DBCO-trimannose targeting moiety

Yeast poly(A) polymerase and its buffer (5x) were purchased from Thermo Fisher Scientific, 3'-azido-2',3'-ddATP was purchased from Jena Bioscience. The 3'- end labelling was performed in reference to Croce *et al.*<sup>[58]</sup> The reaction was set up in a 200 µL PCR vial, whereas 9 µL RNase free H<sub>2</sub>O, 2 µL of 3'-azido-2',3'-ddATP (10 mM, 20 mmol, 500 equiv.), 4 µL ORN1 (10 µM, 40 pmol, 1 equiv., purchased from Ella Biotech), 4 µL γPAP buffer (5X) and 1 µL γPAP (600 U/µL) were combined and mixed by pipetting. The samples were incubated for 2 h at 37°C in a T100 Thermo Cycler from Bio-Rad and subsequently cooled to 4 °C for at least 3 min before being purified *via* spin-column purification.

For the SPAAC of the 3'-N<sub>3</sub> labelled ORN1 (ORN1-a) with the DBCO- trimannose targeting ligand (10 mM stock solution in DMSO, provided by colleagues from the baseclick GmbH), different conditions were tested, as displayed in Table 9.

**Table 9.** SPAAC conditions.

condition	ratio	time [h]	temperature
	ORN1-a: DBCO-triman.		
A1	1:100	3	r.t.
A2	1:100	3	r.t.
B1	1:1000	18	r.t.
B2	1:1000	18	r.t.
B3	1:1000	1	37°

All click reactions were prepared in 0.2 mL PCR tubes. ORN1-a (80 ng in H<sub>2</sub>O, 12 pmol, 1 equiv.) from the previous step was mixed with either 1.2  $\mu$ L (1.41  $\mu$ g, 1200 pmol, 100 equiv.) or 12  $\mu$ L (14.1  $\mu$ g, 12000 pmol, 1000 equiv.) DBCO-trimannose ligand (MW: 1174.21 g/mol, 1 mM dilution of 10 mM stock solution) and filled up with RNase free H<sub>2</sub>O to a total volume of 20  $\mu$ L. Depending on the conditions the reaction mixture was either stored at r.t. for 3 h and 18 h, or incubated at 37°C on a thermocycler for 1 h, before being purified *via* spin column. MALDI-TOF MS analysis were performed as described in section 6.1. and the observed fragments are displayed in Table 10.

**Table 10.** MALDI-TOF MS results

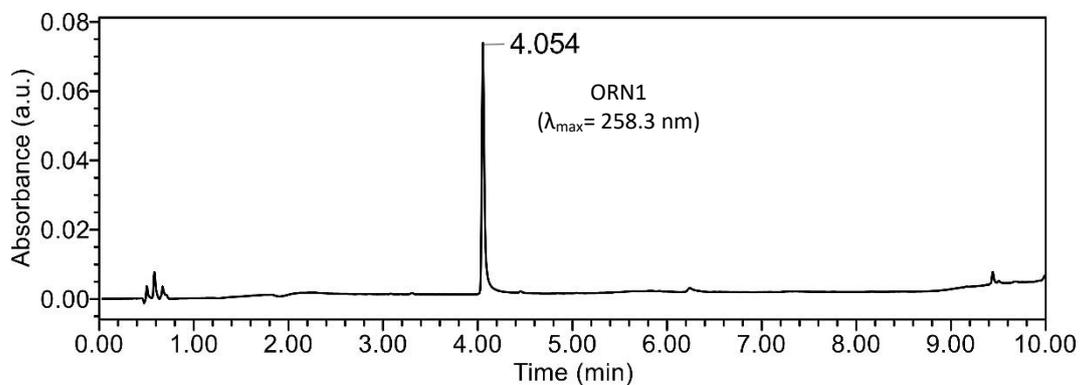
name	sequence (5' to 3')	[M+H] <sup>+</sup> m/z calcd.	[M+H] <sup>+</sup> m/z found
ORN1-a	UUG GUA UCG CUA UCG CUA UGG A-N <sub>3</sub>	7005.20	7006.43
ORN1-b	UUG GUA UCG CUA UCG CUA UGG A-N <sub>3</sub> -DBCO-trimannose conjugate	8178.41	8181.55

The 3'-N<sub>3</sub> labelled of ORN1 and the subsequent SPAAC with the DBCO trimannose ligand using the conditions B2 (1:1000, r.t., 18 h), was further evaluated by analytical RP-HPLC. The results are displayed in Table 11.

**Table 11.** HPL chromatogram of ORN1, ORN1-a and ORN1-b (conditions B2). Gradient: 0-30% buffer B from 0 → 8 min. NSF= no signal found. HPL chromatogram for ORN1-b is only displayed from 2-9 min due to high S/N ratio.

Sequence: **ORN1**

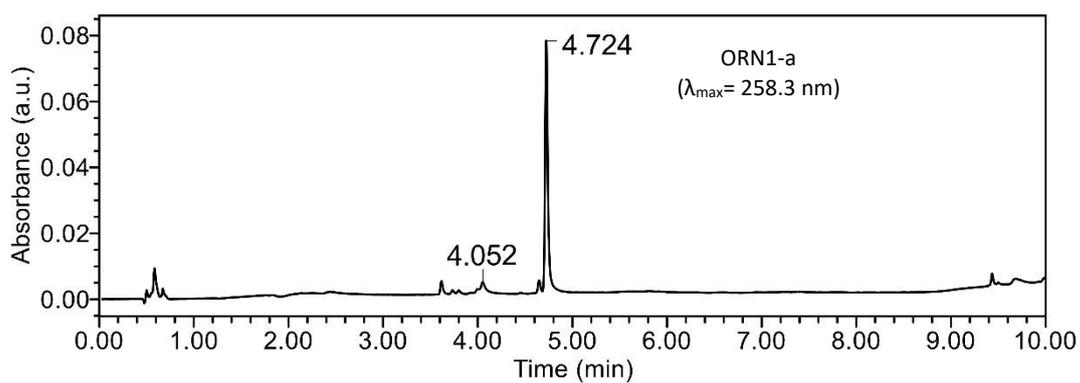
$R_t = 4.054$  min



Sequence: **ORN1-a**

$R_{t(ORN1)} = 4.052$  min,  $R_{t(ORN1-a)} = 4.724$  min

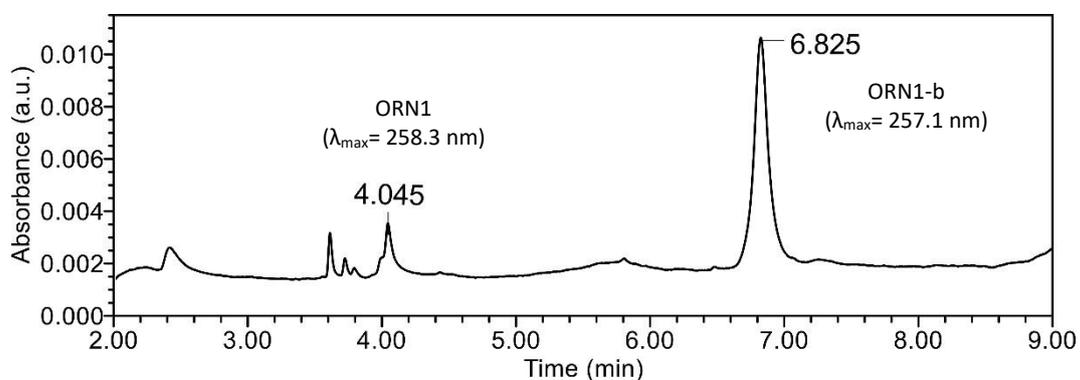
conversion yield: ORN1= 2.3%; ORN1-a=97.7%



Sequence: **ORN1-b**

$R_{t(ORN1)} = 4.045$  min,  $R_{t(ORN1-a)} = \text{NSF}$ ,  $R_{t(ORN1-b)} = 6.825$  min

conversion yield: ORN1=4.8% , ORN1-b=95.2%



The N-protein encoding mRNA with 100% pseudo uridine replacement, was provided by a colleague from the baseclick GmbH. The 3'-N<sub>3</sub> end labelling was performed in a 0.2 mL PCR vial, where 5  $\mu$ L yPAP buffer (5X), N-mRNA (varying volumes, depending on the concentration, 1 equiv.), 1.25  $\mu$ L 3'-azido-2',3'-ddATP (10 mM, 500 equiv.) and 1  $\mu$ L of yPAP (600 U/ $\mu$ L) were combined, filled up to a total volume of 25  $\mu$ L and mixed by pipetting. The samples were incubated for 2 h at 37°C in a T100 Thermal Cycler from Bio-Rad and subsequently cooled to 4 °C for at least 3 min before being purified *via* spin-column purification (RNeasy MinElute Clean up Kit). The SPAAC was performed in a 1:1000 ratio of 3'-N<sub>3</sub> labelled mRNA from the previous step with the DBCO-trimannose ligand (2  $\mu$ L of 10 mM solution), at r.t. for 18 h.

### Quantification by LC-MS

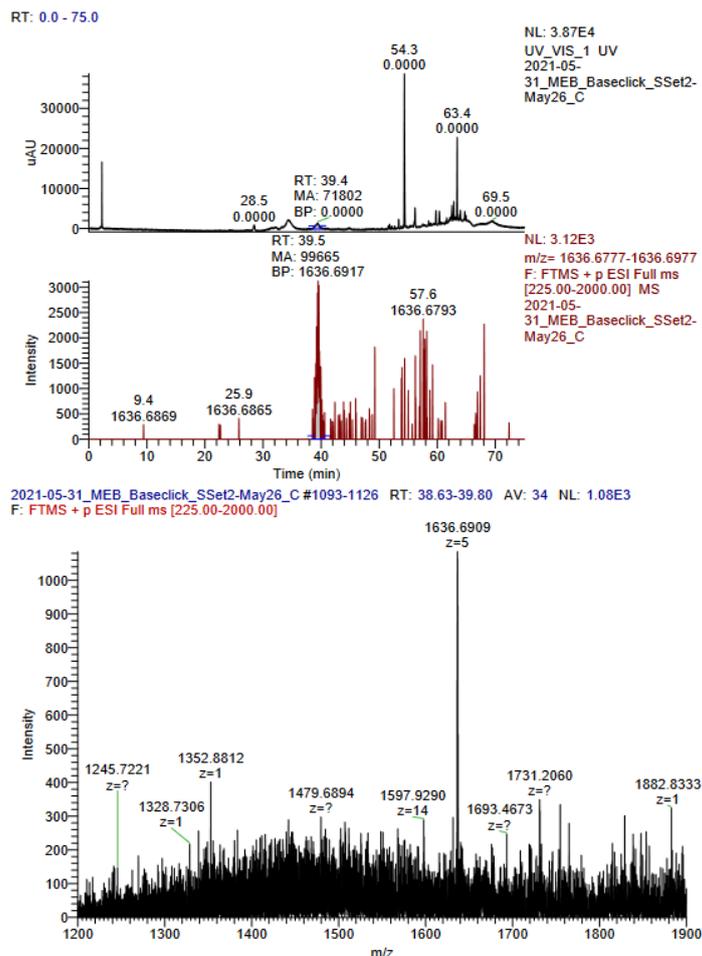
All LC-MS analysis were performed by Dr. Mirko Wagner from the Carell group on a LTQ Orbitrap XL from Thermo Fisher Scientific.

For the LC-MS analysis of the undigested ORN1 samples, ~ 55 ng of each spin column purified sample (ORN1, ORN1-a, ORN1-b) in 100  $\mu$ L RNase free H<sub>2</sub>O, was prepared and stored at -20°C until the LC-MS analysis were performed. The analyzed fragments can be taken from Table 12.

Table 12.

sample	MW [g/mol]	z	MW [g/mol] protonated	[M+H] <sup>+</sup> m/z calcd.
ORN1	6666.9742	4	6671.0040	1667.4725
ORN1-a	7005.1980	4	7009.2278	1751.9885
ORN1-b	8179.4080	5	8183.4378	1636.6876

One exemplary LC-MS result is displayed in Figure 46.



**Figure 46.** LC-MS result of ORN1-b using SPAAC conditions A1 (1:100, 3 h, r.t.). RT= retention time, MA= mass area (mass trace was used for quantification), BP= base peak.

The determined yields are displayed in Table 13.

**Table 13.** Conditions and yield of the identified products ORN1-a and ORN1-b [%]. NSF= no signal found.

condition	ORN1-a [%]	ORN1-b [%]
A1	68.24	31.76
A2	54.42	45.58
B1	9.86	90.14
B2	NSF	100
B3	56.28	43.72

For the LC-MS analysis of the nucleoside digestion products, the all samples were incubated with the NEB Nucleoside Digestion Mix. For the ORN1 samples, 10  $\mu$ L Nucleoside Digestion Mix reaction buffer (10X), between 8 and 12  $\mu$ L of oligoribonucleotide sample (60-100 ng of ORN1, ORN1-a and ORN1-b respectively) and 2  $\mu$ L Nucleoside Digestion Mix were added to a

0.2  $\mu\text{L}$  PCR vial, filled up with RNase free  $\text{H}_2\text{O}$  a total volume of 100  $\mu\text{L}$  and mixed by pipetting. For the mRNA samples, 10  $\mu\text{L}$  of Nucleoside Digestion Mix reaction buffer (10X), 6  $\mu\text{L}$  of mRNA (1  $\mu\text{g}$  of unmodified, 3'- $\text{N}_3$ -labelled and click N-protein encoding mRNA respectively), 2  $\mu\text{L}$  Nucleoside Digestion Mix and 81  $\mu\text{L}$  RNase free  $\text{H}_2\text{O}$  to reach a total volume of 100  $\mu\text{L}$  were added to a 0.2  $\mu\text{L}$  PCR vial and mixed by pipetting. All samples were incubated for 3 h at 37°C and after cooling to r.t. stored at -20°C until the LC-MS analysis were performed.

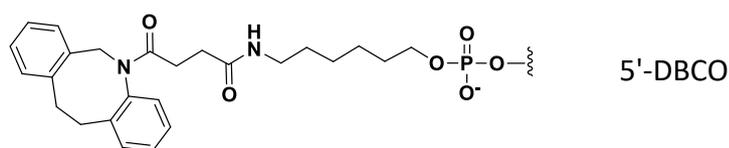
## 6.5.2 Circularization of an RNA-RNA Oligoribonucleotide Assembly *via* SPAAC and Enzymatic Ligation for Potential Circular mRNA Application

The oligoribonucleotides ORN1 and ORN2 were purchased from Ella Biotech GmbH and the sequences of all relevant oligoribonucleotides are displayed in Table 14.

**Table 14.** Sequences of ORN1 and ORN2.

name	type	sequence (5' to 3')	modification	length (nt)	[M+H] <sup>+</sup> m/z calcd.
ORN1	RNA	UUG GUA UCG CUA UCG CUA UGG	-	21	6666.97
ORN1-a	modified RNA	UUG GUA UCG CUA UCG CUA UGG A-N <sub>3</sub>	3' N <sub>3</sub>	22	7005.20
ORN2	modified RNA	YGU AUC GCU AUC GCU A	Y= 5'-DBCO* 3' = phosphate	15	5258.78

\*



**3'-N<sub>3</sub>-labelling *via* yPAP and 3'-Azido-2',3'-ddATP.** The reaction was performed as described in section 6.5.1. MALDI-TOF MS analysis were performed as described in section 6.1.

ORN1-a: [M+H]<sup>+</sup> m/z calcd. 7005.20

found 7004.92

**SPAAC of ORN1-a and ORN2.** To a 0.2 mL PCR vial, 6 μL HEPES (0.1 M in RNase free H<sub>2</sub>O, pH 7-7.6), 15.8 μL NaCl (0.2 M in RNase free H<sub>2</sub>O), 1.4 μL EDTA (0.1 M in RNase free H<sub>2</sub>O), 0.8 μL of ORN2 (421 ng, 80 pmol, 2.5 equiv., 100 μM in RNase free H<sub>2</sub>O) and ORN1-a from the previous step (224 ng, 32 pmol, varying volume depending on the concentration, ~20 ng/μL) were added and the reaction mixture was filled up with RNase free H<sub>2</sub>O to a total volume of 37 μL and mixed by vortexing. The reaction was performed at r.t over night (ca. 18 h), before being purified *via* spin column. **Ligation of ORN1-b *via* RtcB ligase.** The enzymatic ligation was performed in reference to the manufacturer's protocol with adjustments. RtcB ligase, its buffer, MnCl<sub>2</sub>, GTP were purchased from New England Biolabs. To a 0.2 mL PCR vial, 2 μL of RtcB buffer (10X), 2 μL MnCl<sub>2</sub> (10 mM), 2 μL GTP (1 mM), ORN1- b from the previous step

(10 pmol, 119 ng, varying volume depending on the concentration,  $\sim 20$  ng/ $\mu$ L) and 1  $\mu$ L of RtcB ligase (15  $\mu$ M) were added, filled up with RNase free H<sub>2</sub>O to a total volume of 20  $\mu$ L and mixed by pipetting. The reaction mixture was incubated for 2 h at 37°C in a thermocycler and cooled to 4°C before being purified *via* spin column.

**Digestion of ORN1-c *via* RNase R.** The digestion with RNase R (purchased from Biozym/Lucigen) was performed in reference to the manufacturer's protocol with slight adjustments. To a 0.2 mL PCR vial, 2  $\mu$ L RNase R buffer (10X, purchased from Biozym/Lucigen), 0.5  $\mu$ L RNase inhibitor-recombinant (40 U/ $\mu$ L, purchased from Jena Bioscience), ORN1-c (10 pmol, varying volume depending on the concentration,  $\sim 20$  ng/ $\mu$ L) and 2  $\mu$ L of RNase R (10 U/ $\mu$ L) were added, filled up with RNase free H<sub>2</sub>O to a total volume of 20  $\mu$ L and mixed by pipetting. The reaction mixture was incubated for 2 h at 37°C in a thermocycler and cooled to 4°C before being purified *via* spin column.

**Simultaneous SPAAC and ligation of ORN1-a and ORN1 with PEG6000.** To a 0.2 mL PCR vial, 2  $\mu$ L RtcB buffer (10X), 2  $\mu$ L MnCl<sub>2</sub> (10 mM), 2  $\mu$ L GTP (1 mM), 6.25  $\mu$ L or 9.8  $\mu$ L of PEG 6000 (24% solution in H<sub>2</sub>O, to reach a total concentration of 7.5% or 11.8%), ORN1-a (10 pmol, 1 equiv., varying volume depending on the concentration,  $\sim 20$  ng/ $\mu$ L), 1  $\mu$ L ORN2 (10 pmol, 1 equiv., 10 mM in H<sub>2</sub>O) and 1  $\mu$ L of RtcB ligase (15  $\mu$ M) were added, filled up with RNase free H<sub>2</sub>O to a total volume of 20  $\mu$ L and mixed by pipetting. The reaction mixture was incubated for 2 h at 37°C in a thermocycler and cooled to 4°C before being purified *via* spin column.

**dPAGE.** The samples were analyzed by dPAGE as described in section 6.1. Apart from the RNase R treated samples, always 40 ng of each sample were loaded onto the gel. The RNase R treated samples were, after spin column purification, concentrated and loaded entirely onto the gel to ensure comparable results of the actual sample and the control sample. Before being loaded onto the gels, all samples were heat denatured at 65°C for 5 min and then quickly cooled to 4°C. The gels were run at a constant voltage of 150 V for 1.25 h.

## 7. References

- [1] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2001**, *40*, 2004-2021.
- [2] C. W. Tornøe, C. Christensen, M. Meldal, *J. Org. Chem.* **2002**, *67*, 3057-3064.
- [3] N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J. Am. Chem. Soc.* **2004**, *126*, 15046-15047.
- [4] H. L. Davide Castelvechi, Chemists who invented revolutionary 'click' reactions win Nobel, <https://www.nature.com/articles/d41586-022-03087-8#ref-CR4>
- [5] J. E. Moses, A. D. Moorhouse, *Chem. Soc. Rev.* **2007**, *36*, 1249-1262.
- [6] R. Huisgen, *Angew. Chem. Int. Ed.* **1963**, *2*, 565-598.
- [7] L. J. Wang, Y. Tang, in *Comprehensive Organic Synthesis (Second Edition)* (Ed.: P. Knochel), Elsevier, Amsterdam, **2014**, pp. 1342-1383.
- [8] M. Breugst, H.-U. Reissig, *Angew. Chem. Int. Ed.* **2020**, *59*, 12293-12307.
- [9] N. Z. Fantoni, A. H. El-Sagheer, T. Brown, *Chem. Rev.* **2021**, *121*, 7122-7154.
- [10] V. V. Rostovtsev, L. G. Green, V. V. Fokin, et al., *Angew. Chem. Int. Ed.* **2002**, *41*, 2596-2599.
- [11] F. Himo, T. Lovell, R. Hilgraf, et al., *J. Am. Chem. Soc.* **2005**, *127*, 210-216.
- [12] R. Berg, B. F. Straub, *Beilstein J. Org. Chem.* **2013**, *9*, 2715-2750.
- [13] B. T. Worrell, J. A. Malik, V. V. Fokin, *Science* **2013**, *340*, 457-460.
- [14] E. M. Sletten, C. R. Bertozzi, *Acc. Chem. Res.* **2011**, *44*, 666-676.
- [15] R. E. Bird, S. A. Lemmel, X. Yu, et al., *Bioconjugate Chem.* **2021**, *32*, 2457-2479.
- [16] T. R. Chan, R. Hilgraf, K. B. Sharpless, et al., *Org. Lett.* **2004**, *6*, 2853-2855.
- [17] S. I. Presolski, V. Hong, S.-H. Cho, et al., *J. Am. Chem. Soc.* **2010**, *132*, 14570-14576.
- [18] S. L. Scinto, D. A. Bilodeau, R. Hincapie, et al., *Nat. Rev. Methods Primers* **2021**, *1*, 30.
- [19] C. Besanceney-Webler, H. Jiang, T. Zheng, et al., *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 8051-8056.
- [20] M. Meldal, C. W. Tornøe, *Chem. Rev.* **2008**, *108*, 2952-3015.
- [21] J. Kallick, S. Harris, A. K. Udit, et al., *BioTechniques* **2015**, *59*, 329-334.
- [22] B. C. Boren, S. Narayan, L. K. Rasmussen, et al., *J. Am. Chem. Soc.* **2008**, *130*, 8923-8930.
- [23] J. R. Johansson, T. Beke-Somfai, A. Said Stålsmeden, et al., *Chem. Rev.* **2016**, *116*, 14726-14768.
- [24] S. Epple, A. Modi, Y. R. Baker, et al., *J. Am. Chem. Soc.* **2021**, *143*, 16293-16301.

- [25] J. Dommerholt, F. P. J. T. Rutjes, F. L. van Delft, *Top. Curr. Chem.* **2016**, *374*, 16.
- [26] L.-H. Qin, W. Hu, Y.-Q. Long, *Tetrahedron Lett.* **2018**, *59*, 2214-2228.
- [27] J. Dommerholt, O. van Rooijen, A. Borrmann, et al., *Nat. Commun.* **2014**, *5*, 5378.
- [28] M. L. Blackman, M. Royzen, J. M. Fox, *J. Am. Chem. Soc.* **2008**, *130*, 13518-13519.
- [29] F. Thalhammer, U. Wallfahrer, J. Sauer, *Tetrahedron Lett.* **1990**, *31*, 6851-6854.
- [30] D. M. Patterson, L. A. Nazarova, J. A. Prescher, *ACS Chem. Biol.* **2014**, *9*, 592-605.
- [31] D. M. Patterson, L. A. Nazarova, B. Xie, et al., *J. Am. Chem. Soc.* **2012**, *134*, 18638-18643.
- [32] B. L. Oliveira, Z. Guo, O. Boutureira, et al., *Angew. Chem. Int. Ed. Engl.* **2016**, *55*, 14683-14687.
- [33] D. Ganz, P. Geng, H.-A. Wagenknecht, *ACS Chem. Biol.* **2023**, *18*, 1054-1059.
- [34] B. L. Oliveira, Z. Guo, G. J. L. Bernardes, *Chem. Soc. Rev.* **2017**, *46*, 4895-4950.
- [35] ATD Bio, DNA Click chemistry, 19.11.2023, <https://atdbio.com/nucleic-acids-book/Click-chemistry-and-nucleic-acids2023>
- [36] G. T. Hermanson, in *Bioconjugate Techniques (Third Edition)* (Ed.: G. T. Hermanson), Academic Press, Boston, **2013**, pp. 229-258.
- [37] G. S. Liou, H. J. Yen, in *Polymer Science: A Comprehensive Reference* (Eds.: K. Matyjaszewski, M. Möller), Elsevier, Amsterdam, **2012**, pp. 497-535.
- [38] J. Gierlich, G. A. Burley, P. M. E. Gramlich, et al., *Org. Lett.* **2006**, *8*, 3639-3642.
- [39] J. Gierlich, K. Gutsmedl, P. M. E. Gramlich, et al., *Chem. Eur. J.* **2007**, *13*, 9486-9494.
- [40] GlenResearch, Modifier Phosphoramidites and Supports, 19.11.2023, <https://www.glenresearch.com/browse/modifier-phosphoramidites-and-supports2023>
- [41] F. Seela, V. R. Sirivolu, *Chem. Biodivers.* **2006**, *3*, 509-514.
- [42] ATD Bio, Solid-phase oligonucleotide synthesis, 20.11.2023, <https://atdbio.com/nucleic-acids-book/Solid-phase-oligonucleotide-synthesis#Detritylation-of-the-support-bound-3-nucleoside2023>
- [43] F. Seela, V. R. Sirivolu, P. Chitpepu, *Bioconjugate Chem.* **2008**, *19*, 211-224.
- [44] S. A. Ingale, H. Mei, P. Leonard, et al., *J. Org. Chem.* **2013**, *78*, 11271-11282.
- [45] M. Slavičková, M. Janoušková, A. Šimonová, et al., *Chem. Eur. J.* **2018**, *24*, 8311-8314.
- [46] P. M. E. Gramlich, S. Warncke, J. Gierlich, et al., *Angew. Chem. Int. Ed.* **2008**, *47*, 3442-3444.

- [47] A. Salic, T. J. Mitchison, *PNAS* **2008**, *105*, 2415-2420.
- [48] A. B. Neef, N. W. Luedtke, *PNAS* **2011**, *108*, 20404-20409.
- [49] E. Paredes, S. R. Das, *ChemBioChem* **2011**, *12*, 125-131.
- [50] Z. G. Chidgeavadze, R. S. Beabealashvilli, A. A. Krayevsky, et al., *Biochim. Biophys. Acta* **1986**, *868*, 145-152.
- [51] R. S. Beabealashvilli, A. V. Scamrov, T. V. Kutateladze, et al., *Biochim. Biophys. Acta* **1986**, *868*, 136-144.
- [52] A. H. El-Sagheer, T. Brown, *J. Am. Chem. Soc.* **2009**, *131*, 3958-3964.
- [53] A. H. El-Sagheer, A. P. Sanzone, R. Gao, et al., *PNAS* **2011**, *108*, 11338-11343.
- [54] A. H. El-Sagheer, T. Brown, *Chem. Commun.* **2011**, *47*, 12057-12058.
- [55] T. N. Prize, The 2023 Nobel Prize in Physiology or Medicine, 13.12.2023, <https://www.nobelprize.org/prizes/medicine/2023/press-release/2023>
- [56] K. Karikó, H. Muramatsu, F. A. Welsh, et al., *Mol. Ther.* **2008**, *16*, 1833-1840.
- [57] O. Andries, S. Mc Cafferty, S. C. De Smedt, et al., *J. Controlled Release* **2015**, *217*, 337-344.
- [58] S. Croce, S. Serdjukow, T. Carell, et al., *ChemBioChem* **2020**, *21*, 1641-1646.
- [59] S. Maassen, B. Coenen, S.-L. Dülk, et al., *ChemBioChem* **2023**, *24*, e202200658.
- [60] K. Astakhova, R. Ray, M. Taskova, et al., *Mol. Pharm.* **2018**, *15*, 2892-2899.
- [61] S. S. Yu, C. M. Lau, W. J. Barham, et al., *Mol. Pharm.* **2013**, *10*, 975-987.
- [62] K. Brunner, J. Harder, T. Halbach, et al., *Angew. Chem. Int. Ed.* **2015**, *54*, 1946-1949.
- [63] J. Willibald, J. Harder, K. Sparrer, et al., *Journal of the American Chemical Society* **2012**, *134*, 12330-12333.
- [64] F. R. Traube, M. Stern, A. J. Tölke, et al., *Angew. Chem. Int. Ed. Engl.* **2022**, *61*, e202204556.
- [65] T. C. Roberts, R. Langer, M. J. A. Wood, *Nat. Rev. Drug Discov.* **2020**, *19*, 673-694.
- [66] A. Astriab-Fisher, M. H. Fisher, R. Juliano, et al., *Biochem. Pharmacol.* **2004**, *68*, 403-407.
- [67] A. H. El-Sagheer, T. Brown, *Chemical Science* **2014**, *5*, 253-259.
- [68] C. Beck, D. Ramanujam, P. Vaccarello, et al., *Nat. Commun.* **2023**, *14*, 4564.
- [69] E. L. van Dijk, Y. Jaszczyszyn, D. Naquin, et al., *Trends Genet.* **2018**, *34*, 666-681.
- [70] A. Routh, P. Ji, E. Jaworski, et al., *Nucleic Acids Res.* **2017**, *45*, e112.

- [71] S. Y. Gene Shen, Lubomir Sebo, Louis Brogley, Modified nucleotide reagents US10676788B2, <https://patents.google.com/patent/US10676788B2/en>
- [72] J. Eid, A. Fehr, J. Gray, et al., *Science* **2009**, *323*, 133-138.
- [73] S. Serdjukow, F. Kink, B. Steigenberger, et al., *Chem. Commun.* **2014**, *50*, 1861-1863.
- [74] D. G. Aruna Ayer, Peter Crisalli, Meng Taing, Tagged multi-nucleotides useful for nucleic acid sequencing WO2017202917A1, <https://patents.google.com/patent/WO2017202917A1/en>
- [75] A. Sood, S. Kumar, S. Nampalli, et al., *J. Am. Chem. Soc.* **2005**, *127*, 2394-2395.
- [76] B. Searle, M. Müller, T. Carell, et al., *Angew. Chem. Int. Ed.* **2023**, *62*, e202215704.
- [77] H. Brinkerhoff, A. S. W. Kang, J. Liu, et al., *Science* **2021**, *374*, 1509-1513.
- [78] E. S. Schönegger, A. Crisp, M. Müller, et al., *Bioconjugate Chem.* **2022**, *33*, 1789-1795.
- [79] R. Stark, M. Grzelak, J. Hadfield, *Nat. Rev. Genet.* **2019**, *20*, 631-656.
- [80] D. A. Shagin, K. A. Lukyanov, L. L. Vagner, et al., *Nucleic Acids Res.* **1999**, *27*, e23-i-e23-iii.
- [81] A. Bayega, S. Oikonomopoulos, Y. C. Wang, et al., *bioRxiv* **2022**, 2022.2007.2028.501892.
- [82] ONT, Direct cDNA sequencing (SQK-LSK114), 21.08.2023, [https://community.nanoporetech.com/docs/prepare/library\\_prep\\_protocols/ligation-sequencing-v14-direct-cdna-sequencing/v/dcs\\_9187\\_v114\\_rev19apr2023/reverse-transcription-and-strand-switching?devices=minion](https://community.nanoporetech.com/docs/prepare/library_prep_protocols/ligation-sequencing-v14-direct-cdna-sequencing/v/dcs_9187_v114_rev19apr2023/reverse-transcription-and-strand-switching?devices=minion)
- [83] ONT, cDNA-PCR Sequencing (SQK-PCS111), 21.08.2023, [https://community.nanoporetech.com/docs/prepare/library\\_prep\\_protocols/pcr-cdna-sequencing-sqk-pcs111/v/pcs\\_9143\\_v111\\_revo\\_01dec2021/reverse-transcription-and-strand-switching?devices=minion](https://community.nanoporetech.com/docs/prepare/library_prep_protocols/pcr-cdna-sequencing-sqk-pcs111/v/pcs_9143_v111_revo_01dec2021/reverse-transcription-and-strand-switching?devices=minion)
- [84] K. Athanopoulou, M. A. Boti, P. G. Adamopoulos, et al., *Life* **2022**, *12*, 30.
- [85] S. Picelli, O. R. Faridani, Å. K. Björklund, et al., *Nature Protocols* **2014**, *9*, 171-181.
- [86] Y. Wang, Y. Zhao, A. Bollas, et al., *Nat. Biotechnol.* **2021**, *39*, 1348-1365.
- [87] E. S. Schönegger, A. Crisp, M. Radukic, et al., *ChemBioChem* **2023**, *n/a*, e202300701.
- [88] A. Ramanathan, G. B. Robb, S.-H. Chan, *Nucleic Acids Res.* **2016**, *44*, 7511-7526.
- [89] M. Shanmugasundaram, A. Senthilvelan, A. R. Kore, *The Chemical Record* **2022**, *22*, e202200005.

- [90] J. Jemielity, T. Fowler, J. Zuberek, et al., *RNA* **2003**, *9*, 1108-1122.
- [91] E. Fang, X. Liu, M. Li, et al., *Signal Transduction and Targeted Therapy* **2022**, *7*, 94.
- [92] J. Stepinski, C. Waddell, R. Stolarski, et al., *RNA* **2001**, *7*, 1486-1495.
- [93] J. M. Henderson, A. Ujita, E. Hill, et al., *Current Protocols* **2021**, *1*, e39.
- [94] P. M. D. Moreno, M. Wenska, K. E. Lundin, et al., *Nucleic Acids Res.* **2009**, *37*, 1925-1935.
- [95] B. A. Wojtczak, M. Warminski, J. Kowalska, et al., *RSC Advances* **2016**, *6*, 8317-8328.
- [96] M. Honcharenko, J. Romanowska, M. Alvira, et al., *RSC Advances* **2012**, *2*, 12949-12962.
- [97] T. M. Anthony, M. K. H. Pflum, *Biorg. Med. Chem.* **2018**, *26*, 2331-2336.
- [98] S. M. Hacker, M. Mex, A. Marx, *J. Am. Chem. Soc.* **2012**, *77*, 10450-10454.
- [99] M. L. Winz, A. Samanta, D. Benzinger, et al., *Nucleic Acids Res.* **2012**, *40*, e78.
- [100] S. M. Hacker, M. Welter, A. Marx, *Curr. Protoc. Nucleic Acid Chem.* **2015**, *60*, 13.14.11-13.14.25.
- [101] K. Kettenbach, T. L. Ross, *MedChemComm* **2016**, *7*, 654-657.
- [102] G. F. Chimonides, A. A. Sohdi, M. R. Khaleghi, et al., *J. Polym. Sci., Part A: Polym. Chem.* **2013**, *51*, 4853-4859.
- [103] U. Rieder, N. W. Luedtke, *Angew. Chem. Int. Ed.* **2014**, *53*, 9168-9172.
- [104] B. L. Oliveira, Z. Guo, O. Boutureira, et al., *Angew. Chem. Int. Ed.* **2016**, *55*, 14683-14687.
- [105] E. A. Galburt, J. Pelletier, G. Wilson, et al., *Structure* **2002**, *10*, 1249-1260.
- [106] Jena Bioscience, Pyrimidyl-Tetrazine-Cy3, 17.12.2023,  
<https://www.jenabioscience.com/click-chemistry/click-reagents-by-chemistry/tetrazine-reagents/fluorescent-dyes/clk-100-pyrimidyl-tetrazine-cy32023>
- [107] S. Eising, B.-T. Xin, F. Kleinpenning, et al., *ChemBioChem* **2018**, *19*, 1648-1652.
- [108] C. E. Hoyle, C. N. Bowman, *Angew. Chem. Int. Ed.* **2010**, *49*, 1540-1573.
- [109] W. Xi, S. Pattanayak, C. Wang, et al., *Angew. Chem. Int. Ed.* **2015**, *54*, 14462-14467.
- [110] G. A. Logsdon, M. R. Vollger, E. E. Eichler, *Nat. Rev. Genet.* **2020**, *21*, 597-614.
- [111] S. Ermert, S. M. Hacker, A. Buntru, et al., *ChemBioChem* **2017**, *18*, 378-381.
- [112] S. Ermert, A. Marx, S. M. Hacker, *Top Curr Chem (Cham)* **2017**, *375*, 28.
- [113] S. E. Lee, L. M. Elphick, A. A. Anderson, et al., *Bioorg Med Chem Lett* **2009**, *19*, 3804-3807.

- [114] S. Zhang, N. Zhang, J. C. Blain, et al., *J. Am. Chem. Soc.* **2013**, *135*, 924-932.
- [115] M. Rachwalak, J. Romanowska, M. Sobkowski, et al., *Appl. Sci.* **2021**, *11*, 2248.
- [116] N. Zhu, D. Zhang, W. Wang, et al., *N Engl J Med* **2020**, *382*, 727-733.
- [117] F. Wu, S. Zhao, B. Yu, et al., *Nature* **2020**, *579*, 265-269.
- [118] E. Petersen, M. Koopmans, U. Go, et al., *Lancet Infect. Dis.* **2020**, *20*, e238-e244.
- [119] WHO, Coronavirus disease (COVID-19)-Fact sheets, 06.09.2023, [https://www.who.int/news-room/fact-sheets/detail/coronavirus-disease-\(covid-19\)](https://www.who.int/news-room/fact-sheets/detail/coronavirus-disease-(covid-19))
- [120] O. J. Wouters, K. C. Shadlen, M. Salcher-Konrad, et al., *Lancet* **2021**, *397*, 1023-1034.
- [121] A. Kuzmina, Y. Khalaila, O. Voloshin, et al., *Cell Host Microbe* **2021**, *29*, 522-528.e522.
- [122] C. Zhu, J. Y. Lee, J. Z. Woo, et al., *Nat. Commun.* **2022**, *13*, 4503.
- [123] Y. Huang, C. Yang, X. F. Xu, et al., *Acta Pharmacol Sin* **2020**, *41*, 1141-1149.
- [124] L. Wang, L. Wang, H. Zhuang, *Signal Transduct. Target. Ther.* **2020**, *5*, 185.
- [125] N. G. Davies, C. I. Jarvis, K. van Zandvoort, et al., *Nature* **2021**, *593*, 270-274.
- [126] M. M. Lamers, B. L. Haagmans, *Nat. Rev. Microbiol.* **2022**, *20*, 270-284.
- [127] WHO, Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic, [https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-coronavirus-disease-\(covid-19\)-pandemic](https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic)
- [128] S. Hegde, Z. Tang, J. Zhao, et al., *Front. Chem.* **2022**, *10*.
- [129] R. Lu, Zhao, X., Li, J., Niu, P., Yang, B, Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., X. Ma, Zhan, F., Wang, L., Hu, T., H. Zhou, et al., *Lancet* **2020**, *395*, 565-574.
- [130] S. Alexandersen, A. Chamings, T. R. Bhatta, *Nat. Commun.* **2020**, *11*, 6059.
- [131] N. K. Dutta, K. Mazumdar, J. T. Gordy, *J. Virol.* **2020**, *94*.
- [132] J. Shang, G. Ye, K. Shi, et al., *Nature* **2020**, *581*, 221-224.
- [133] C. B. Jackson, M. Farzan, B. Chen, et al., *Nat. Rev. Mol. Cell Biol.* **2022**, *23*, 3-20.
- [134] W. Wu, Y. Cheng, H. Zhou, et al., *Virol. J.* **2023**, *20*, 6.
- [135] C.-k. Chang, M.-H. Hou, C.-F. Chang, et al., *Antiviral Res.* **2014**, *103*, 39-50.
- [136] Z. Zhang, N. Nomura, Y. Muramoto, et al., *Nat. Commun.* **2022**, *13*, 4399.
- [137] Y. L. Siu, K. T. Teoh, J. Lo, et al., *J. Virol.* **2008**, *82*, 11318-11330.
- [138] S. Zhou, P. Lv, M. Li, et al., *Biomed Pharmacother* **2023**, *159*, 114242.
- [139] C. Liu, L. Mendonça, Y. Yang, et al., *Structure* **2020**, *28*, 1218-1224.e1214.

- [140] WHO, Vaccines Granted Emergency Use Listing (EUL) by WHO, <https://covid19.trackvaccines.org/agency/who/>
- [141] T. K. Le, C. Paris, K. S. Khan, et al., *Trends Biochem. Sci.* **2021**, *46*, 351-365.
- [142] C. Yi, Y. Yi, J. Li, *Virolog. Sin.* **2020**, *35*, 259-262.
- [143] N. Bellamkonda, U. P. Lambe, S. Sawant, et al., *Biomedicines* **2022**, *10*.
- [144] X. Su, W. Ma, D. Feng, et al., *Angew. Chem. Int. Ed.* **2021**, *60*, 21662-21667.
- [145] Y. Qiao, J. W. Wotring, C. J. Zhang, et al., *PLoS One* **2023**, *18*, e0281281.
- [146] P. H. Hagedorn, R. Persson, E. D. Funder, et al., *Drug Discov. Today* **2018**, *23*, 101-114.
- [147] J. C. François, T. Saison-Behmoaras, C. Barbier, et al., *PNAS* **1989**, *86*, 9702-9706.
- [148] T. Niittymäki, H. Lönnberg, *Org. Biomol. Chem.* **2006**, *4*, 15-25.
- [149] A. Kuzuya, M. Komiyama, *Curr. Org. Chem.* **2007**, *11*, 1450-1459.
- [150] O. Luige, P. P. Bose, R. Stulz, et al., *Chem. Commun.* **2021**, *57*, 10911-10914.
- [151] T. Lauria, C. Slator, V. McKee, et al., *Chem. Eur. J.* **2020**, *26*, 16782-16792.
- [152] L. M. Gaetke, C. K. Chow, *Toxicology* **2003**, *189*, 147-163.
- [153] I. Fridovich, *J. Biol. Chem.* **1997**, *272*, 18515-18517.
- [154] B. Halliwell, J. M. C. Gutteridge, *Free Radicals in Biology and Medicine*, Oxford University Press, **2015**.
- [155] Z. Molphy, D. Montagner, S. S. Bhat, et al., *Nucleic Acids Res.* **2018**, *46*, 9918-9931.
- [156] J. P. Kehrer, *Toxicology* **2000**, *149*, 43-50.
- [157] D. R. Lloyd, D. H. Phillips, *Mutat. Res. Fund. Mol. Mech. Mutat.* **1999**, *424*, 23-36.
- [158] C. Slator, Z. Molphy, V. McKee, et al., *Nucleic Acids Res.* **2018**, *46*, 2733-2750.
- [159] N. Zuin Fantoni, B. McGorman, Z. Molphy, et al., *ChemBioChem* **2020**, *21*, 3563-3574.
- [160] B. McGorman, N. Z. Fantoni, S. O'Carroll, et al., *Nucleic Acids Res.* **2022**, *50*, 5467-5481.
- [161] S.-C. Chen, R. C. L. Olsthoorn, *Virology* **2010**, *401*, 29-41.
- [162] Z. Miao, A. Tidu, G. Eriani, et al., *RNA Biol.* **2021**, *18*, 447-456.
- [163] S. P. Ryder, B. R. Morgan, P. Coskun, et al., *Evol. Bioinform.* **2021**, *17*, 11769343211014167.
- [164] V. Lulla, M. P. Wandel, K. J. Bandyra, et al., *J. Virol.* **2021**, *95*, 10.1128/jvi.00663-00621.
- [165] V. M. Farzan, I. O. Aparin, O. A. Veselova, et al., *Anal. Methods* **2016**, *8*, 5826-5831.
- [166] A. R. Gruber, R. Lorenz, S. H. Bernhart, et al., *Nucleic Acids Res.* **2008**, *36*, W70-74.
- [167] P. Kerpedjiev, S. Hammer, I. L. Hofacker, *Bioinform.* **2015**, *31*, 3377-3379.
- [168] V. Hong, S. I. Presolski, C. Ma, et al., *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 9879-9883.

- [169] M. Jerabek-Willemsen, T. André, R. Wanner, et al., *J. Mol. Struct.* **2014**, *1077*, 101-113.
- [170] S. El Deeb, A. Al-Harrasi, A. Khan, et al., *Methods and Applications in Fluorescence* **2022**, *10*, 042001.
- [171] N. T. GmbH, Kd value, 21.12.2023, <https://nanotempertech.com/nanopedia/kd/>
- [172] D. Jacob, K. Thüring, A. Galliot, et al., *Angew. Chem. Int. Ed.* **2019**, *58*, 9565-9569.
- [173] NanoTemper , T. GmbH, Binding curve direction in MicroScale Thermophoresis (MST) 21.12.2023, <https://support.nanotempertech.com/hc/en-us/articles/19814170578449-Binding-curve-direction-in-MicroScale-Thermophoresis-MST->
- [174] R. Stulz, M. Lerche, O. Luige, et al., *RSC Chem. Biol.* **2023**, *4*, 1123-1130.
- [175] A. Panattoni, R. Pohl, M. Hocek, *Org. Lett.* **2018**, *20*, 3962-3965.
- [176] Z. N. Baker, P. A. Cobine, S. C. Leary, *Metallomics* **2017**, *9*, 1501-1512.
- [177] R. B. Franklin, B. Milon, P. Feng, et al., *Front Biosci* **2005**, *10*, 2230-2239.
- [178] D. Beyersmann, *Materialwiss. Werkstofftech.* **2002**, *33*, 764-769.
- [179] N. Al Fayez, M. S. Nassar, A. A. Alshehri, et al., *Pharmaceutics* **2023**, *15*, 1972.
- [180] WHO, **2023**.
- [181] S. A. Dilliard, Q. Cheng, D. J. Siegwart, *PNAS USA* **2021**, *118*.
- [182] T. Liu, Y. Tian, A. Zheng, et al., *Polymers* **2022**, *14*, 4195.
- [183] S. M. Moghimi, D. Simberg, *Mol. Ther.* **2022**, *30*, 2109-2110.
- [184] S. S. T. Frischmuth, B. Graf, S. Croce, Click-modified in vitro transcribed mRNA for gene expression, EP3502258, <https://patentscope.wipo.int/search/en/detail.jsf?docId=EP245432718>
- [185] D. Li, J. Li, *J. Clin. Microbiol.* **2021**, *59*.
- [186] B. Lepenies, J. Lee, S. Sonkaria, *Adv Drug Deliv Rev* **2013**, *65*, 1271-1281.
- [187] K. Wagener, M. Bros, M. Krumb, et al., *Adv. Ther.* **2020**, *3*, 1900185.
- [188] K. Uehara, T. Harumoto, A. Makino, et al., *Nucleic Acids Res.* **2022**, *50*, 4840-4859.
- [189] R. Goswami, D. Chatzikleantous, G. Lou, et al., *ACS Infect. Dis.* **2019**, *5*, 1546-1558.
- [190] M. Beverly, C. Hagen, O. Slack, *Anal. Bioanal. Chem.* **2018**, *410*, 1667-1677.
- [191] R. Chen, S. K. Wang, J. A. Belk, et al., *Nat. Biotechnol.* **2023**, *41*, 262-272.
- [192] C. X. Liu, S. K. Guo, F. Nan, et al., *Mol. Cell* **2022**, *82*, 420-434.e426.
- [193] R. A. Wesselhoeft, P. S. Kowalski, D. G. Anderson, *Nat. Commun.* **2018**, *9*, 2629.
- [194] L. Qu, Z. Yi, Y. Shen, et al., *Cell* **2022**, *185*, 1728-1744.e1716.

- [195] Y. Yang, X. Fan, M. Mao, et al., *Cell Res.* **2017**, *27*, 626-641.
- [196] R. A. Wesselhoeft, P. S. Kowalski, F. C. Parker-Hale, et al., *Mol. Cell* **2019**, *74*, 508-520.e504.
- [197] R. Kumar, A. El-Sagheer, J. Tumpance, et al., *J. Am. Chem. Soc.* **2007**, *129*, 6859-6864.
- [198] M. Warminski, J. Kowalska, J. Jemielity, *Org. Lett.* **2017**, *19*, 3624-3627.
- [199] Q. Kong, C. L. Lin, *Cell. Mol. Life Sci.* **2010**, *67*, 1817-1829.
- [200] A. Curry-Hyde, U. Ueberham, T. Arendt, et al., *Genomics* **2020**, *112*, 1162-1166.
- [201] S. Müller, B. Appel, *RNA Biology* **2017**, *14*, 1018-1027.
- [202] L. Taemaitree, A. Shivalingam, A. H. El-Sagheer, et al., *Nat. Commun.* **2019**, *10*, 1610.
- [203] N. A. Chebotareva, B. I. Kurganov, N. B. Livanova, *Biochem. (Mosc)* **2004**, *69*, 1239-1251.
- [204] X. Li, X. Xiong, K. Wang, et al., *Nat. Chem. Biol.* **2016**, *12*, 311-316.
- [205] V. Shabardina, T. Kischka, F. Manske, et al., *GigaScience* **2019**, *8*.
- [206] H. Li, B. Handsaker, A. Wysoker, et al., *Bioinform.* **2009**, *25*, 2078-2079.
- [207] M. R. Webb, G. P. Reid, V. R. N. Munasinghe, et al., *Biochem.* **2004**, *43*, 14463-14471.

## 8. Appendices

### 8.1 Supporting Information of Published Work

## Supplementary Information

### Click Chemistry enables rapid amplification of full-length reverse transcripts for long read Third Generation Sequencing

Eva S. Schönegger ‡ †, Antony Crisp‡, Markus Müller †, Jessica Fertl, Sascha Serdjukow, Stefano Croce, Michael Kollaschinski, Thomas Carell †, Thomas Frischmuth\*

baseclick GmbH, Floriansbogen 2-4, 82061 Neuried (Munich), Germany

† Ludwig-Maximilians-Universität München, Institute for Chemical Epigenetics Munich, Butenandtstr. 5-13, 81377 Munich, Germany

‡ These authors contributed equally to this work.

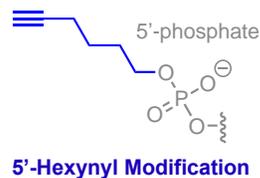
<b>Oligonucleotide sequences</b> .....	2
<b>Agarose gel images and electropherogram data</b> .....	3
<b>Map of the plasmid and sequences of the mRNA model system (eGFP)</b> .....	6
<b>Additional sequencing data analysis and enzyme information</b> .....	7
<b>References</b> .....	9

## Oligonucleotide sequences

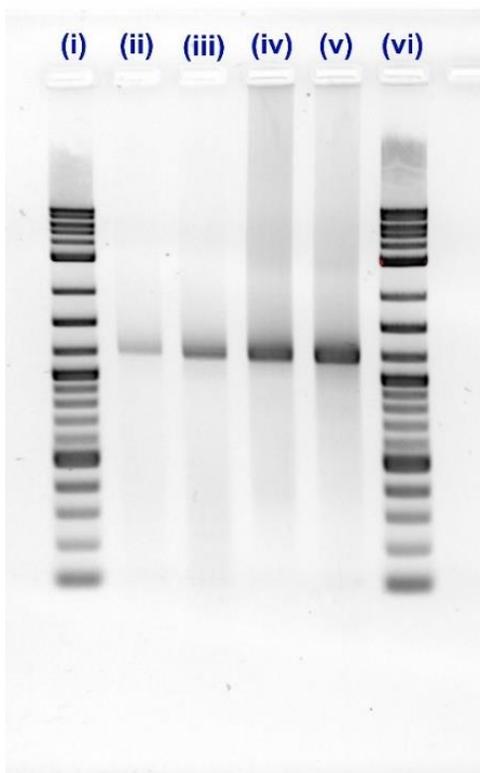
**Table S1.** Synthetic Oligonucleotides and Primers.

Name	Type	Sequence (5' to 3')	Modifications	Usage	Source / Rationale
AA1	Modified-DNA	XTA GAT CCG AAG AGC GTC GTG TAG GGA AAG AGT GTA GAT CTC GGT GGT CGC CGT ATC ATT	X = 5'-Hexynyl-CEP [a]	Alkyne adapter for 3'-end click reaction (CuAAC) with azido-labelled cDNA	Modification of the oligonucleotide "click-adapter" originally designed by Elrod et al. <sup>1</sup>
RTP1	DNA	AAG CAG TGG TAT CAA CGC AGA GTA CTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TVN	V = A, C, G N = A, C, G, T	Reverse Transcription Primer	Modification of the primer "Oligo-dT <sub>30</sub> VN", as described by Picelli et al. <sup>2</sup>
RTP2	DNA	CAA GCA GAA GAC GGC ATA CGA TTT TTT TTT TTT TTT TTT TVN	V = A, C, G N = A, C, G, T	Reverse Transcription Primer	Modification of the primer "MCA02", as described by Han et al. <sup>3</sup>
RTP3	DNA	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TTT TTT TTT TTT TTT TTT TVN	V = A, C, G N = A, C, G, T	Reverse Transcription Primer	The primer "HITS-3", as described by Shepard et al. <sup>4</sup>
RTP4	DNA	ATA TGG ATC CGG CGC GCC GTC GAC TTT TTT TTT TTT TTT TTT TTT TTT VN	V = A, C, G N = A, C, G, T	Reverse Transcription Primer	Modification of the primer "Universal Primer 1", as described by Tang et al. <sup>5</sup>
RTP5	DNA	AAG CAG TGG TAT CAA CGC AGA GTA CTT TTT TTT TTT TTT TTT TTT VN	V = A, C, G N = A, C, G, T	Reverse Transcription Primer	Modification of the primer "3'-RACE CDS primer A", as described by Shi et al. <sup>6</sup>
RTP6	DNA	GTG ACT GGA GTT CAG ACG TGT TTT TTT TTT TTT TTT TTT TVN	V = A, C, G N = A, C, G, T	Reverse Transcription Primer	Modification of the primer "Illumina_4N_21T primer", as described by Elrod et al. <sup>1</sup>
FP1	DNA	[PHO] C GAC GCT CTT CCG ATC TAC	[PHO] = 5'-phosphate	Forward PCR Primer	Complementary to AA1
FP2	DNA	GAC GCT CTT CCG ATC TAC G	n/a	Forward PCR Primer	Complementary to AA1
FP3	DNA	GAC GCT CTT CCG ATC TAC N	N = A, C, G, T	Forward PCR Primer	Complementary to AA1
RP1	DNA	[PHO] AAG CAG TGG TAT CAA CGC AGA GT	[PHO] = 5'-phosphate	Reverse PCR Primer	Complementary to RTP1
RP2	DNA	[PHO] CAA GCA GAA GAC GGC ATA CGA	[PHO] = 5'-phosphate	Reverse PCR Primer	Complementary to RTP2
RP3	DNA	[PHO] TTC CCT ACA CGA CGC TCT TC	[PHO] = 5'-phosphate	Reverse PCR Primer	Complementary to RTP3
RP4	DNA	[PHO] ATA TGG ATC CGG CGC GC	[PHO] = 5'-phosphate	Reverse PCR Primer	Complementary to RTP4
RP5	DNA	[PHO] AAG CAG TGG TAT CAA CGC AG	[PHO] = 5'-phosphate	Reverse PCR Primer	Complementary to RTP5
RP6	DNA	[PHO] GTG ACT GGA GTT CAG ACG TGT	[PHO] = 5'-phosphate	Reverse PCR Primer	Complementary to RTP6

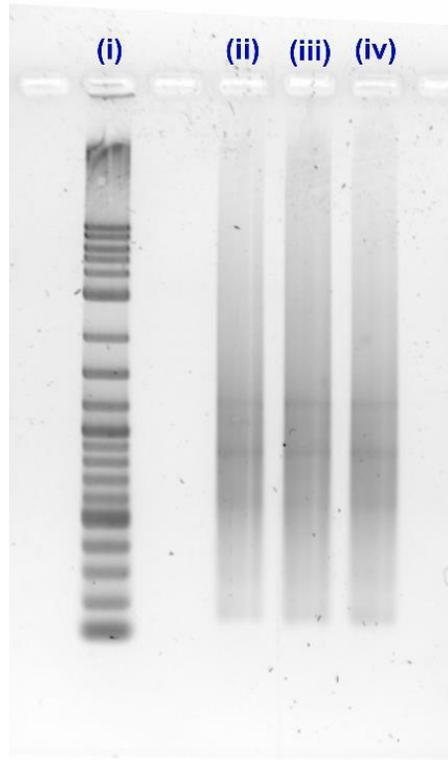
[a] Structure of the modification obtained by after synthesis of an oligonucleotide using 5-Hexyn-1-yl-(2-cyanoethyl)-(N,N-diisopropyl)-phosphoramidite (5'-Hexynyl-CEP) is given below.



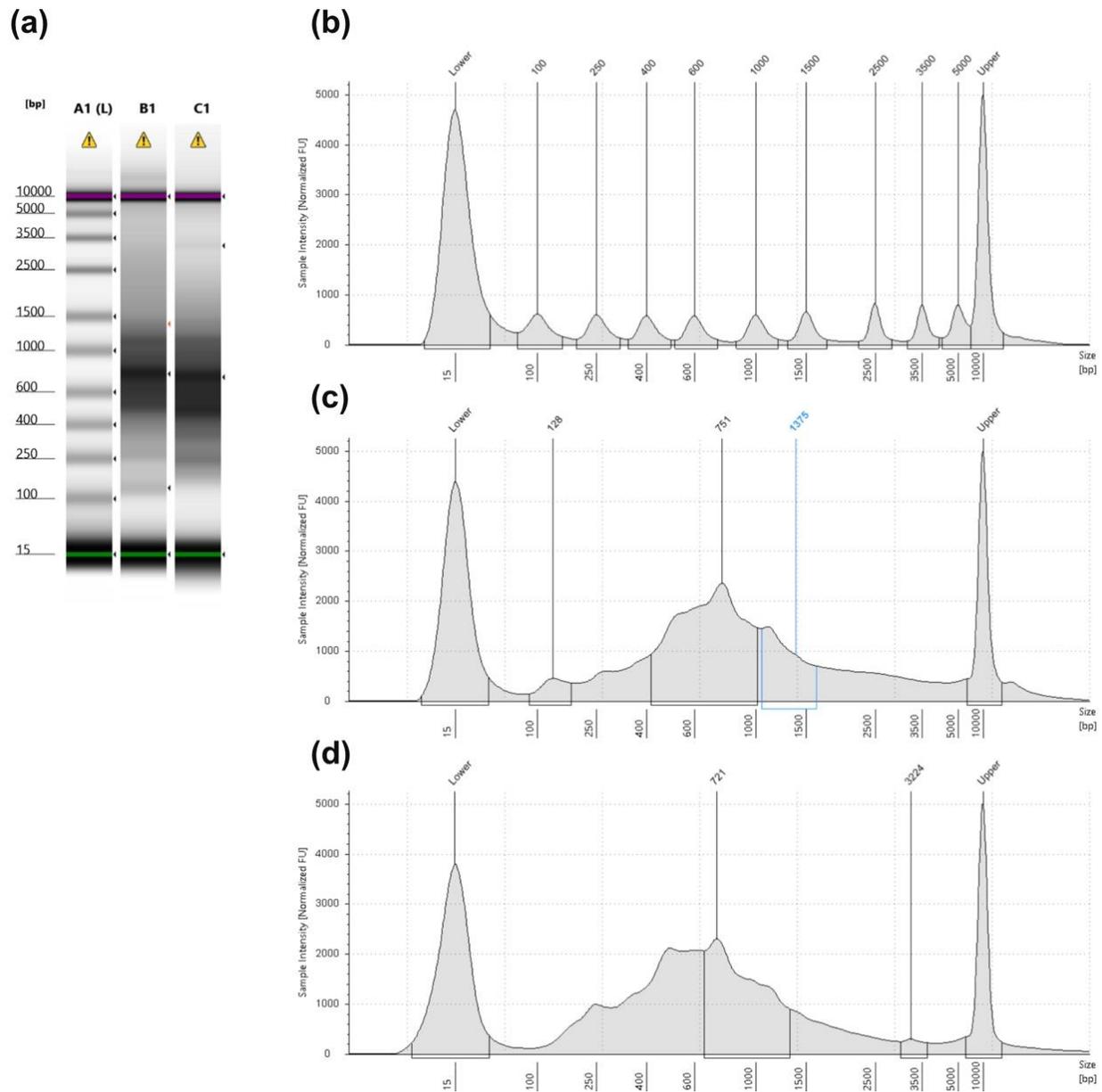
## Agarose gel images and electropherogram data



**Figure S1.** Agarose gel (1.5% w/v in 0.5x TAE buffer) image obtained after electrophoresis (100 V for ca. 1.5 hr) of 1.16 kb amplicons. PCR products were generated by amplification of eGFP-encoding clicked-cDNA with different candidate forward PCR primers. Staining was performed with ethidium bromide. In all cases, RTP3 was employed as the reverse transcription primer, and RP3 as the reverse PCR primer. **(i and vi)** 1 kb Plus DNA Ladder (formerly 2-Log DNA Ladder) from NEW ENGLAND BIOLABS. **(ii)** The PCR product obtained by using FP1 as the forward primer. **(iii)** The PCR product obtained by using FP1 as the forward primer, with twice the previous concentration of DNA loaded onto the gel, as compared with (ii). **(iv)** The PCR product obtained by using FP2 as the forward primer. **(v)** The PCR product obtained by using FP3 as the forward primer.

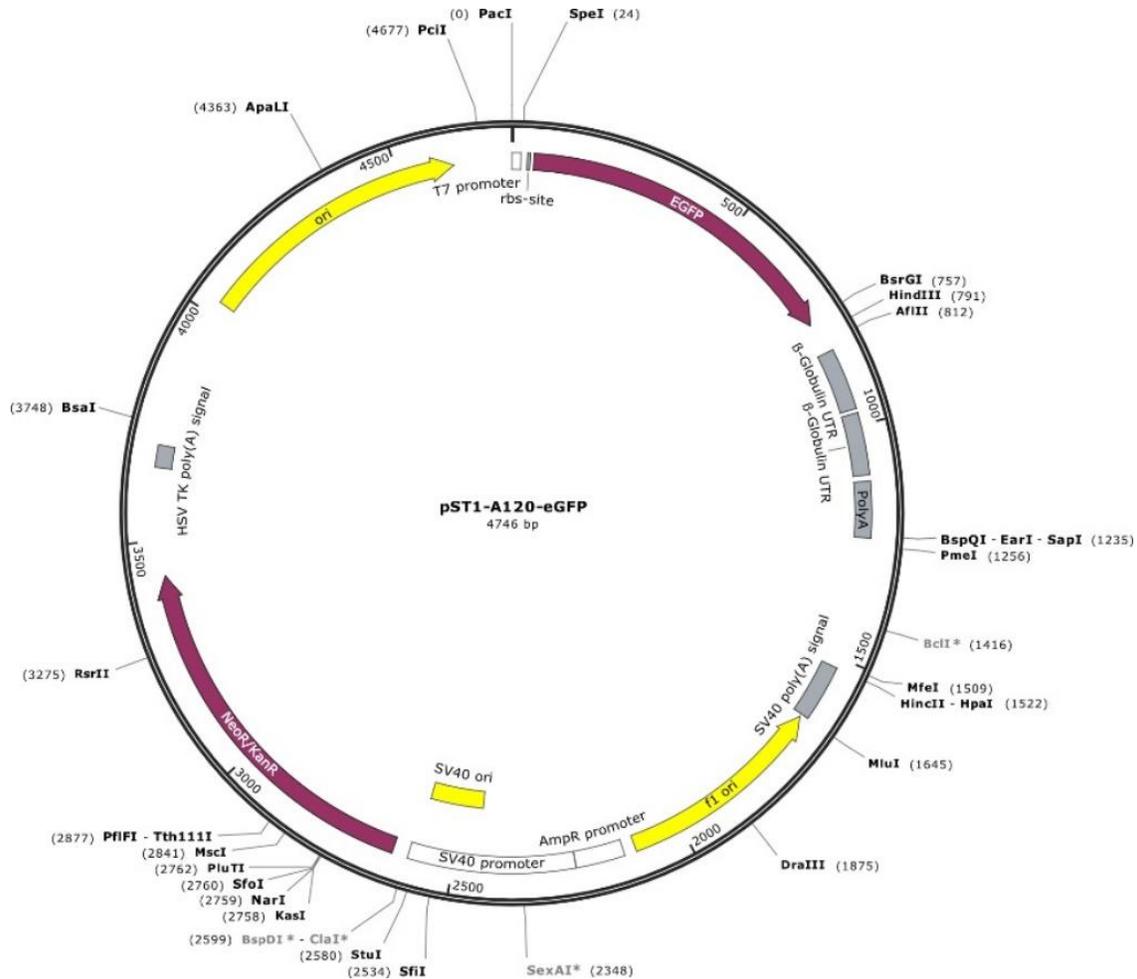


**Figure S2.** Agarose gel (1.5% w/v in 0.5x TAE buffer) image obtained after electrophoresis (100 V for ca. 1.5 hr) of three dsDNA libraries generated from total RNA (3x technical replicates). PCR products were generated by amplification of clicked-cDNA derived from T Cell Leukemia (Jurkat) Total RNA (THERMOFISHER SCIENTIFIC), and subsequently purified *via* spin column. Staining was performed with GelRed® as the dye. **(i)** 1 kb Plus DNA Ladder (formerly 2-Log DNA Ladder) from NEW ENGLAND BIOLABS. **(ii-iv)** dsDNA libraries obtained after reverse transcription using RTP3, followed by PCR with FP3 as the forward primer and RP3 as the reverse primer.



**Figure S3.** Representative electropherogram data of a dsDNA library generated from T Cell Leukemia (Jurkat) Total RNA, before and after preparation for ONT sequencing with a Direct cDNA Sequencing Kit (OXFORD NANOPORE TECHNOLOGIES). PCR products were generated by amplification of clicked-cDNA derived from T Cell Leukemia (Jurkat) Total RNA (THERMO FISHER SCIENTIFIC), and subsequently purified *via* spin column. In all cases, RTP3 was employed as the reverse transcription primer, RP3 as the reverse PCR primer, and FP3 as the forward primer. Each sample was run on a 4200 TapeStation system (AGILENT), equipped with High Sensitivity D5000 ScreenTape. (a) Normalised gel view of the TapeStation run. Sample lanes A1(L), B1, and C1 correspond to the electropherograms (b), (c) and (d) respectively. (b) High Sensitivity D5000 Ladder (AGILENT). (c) The electropherogram measured directly after purification of the PCR product *via* spin column. (d) The electropherogram measured after size selection, end clean-up, and sequencing adaptor ligation using the Direct cDNA Sequencing Kit (OXFORD NANOPORE TECHNOLOGIES).

## Map of the plasmid and sequences of the mRNA model system (eGFP)



**Figure S4.** Map of the plasmid used in this study to generate the eGFP-encoding mRNA by means of T7 RNA polymerase reaction. Relevant sequence segments are labelled as depicted.<sup>7</sup> Graphical visualisation of the annotated plasmid map was generated and exported using the software SnapGene (DOTMatics).

5' -GGGCGAACTAGTAAGCAAGGAGGCGTGCAGATGTGAGCAAGGGCGAGGAGCTGTTACCCGGGGTGGTGCCCATCCTGGT  
 CGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC  
 TGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCTGGCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTC  
 AGCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTT  
 CTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGG  
 GCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATCATGGCC  
 GACAAGCAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTA  
 CCAGCAGAACACCCCATCGGGCAGCGCCCGTGTCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAG  
 ACCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTGGCATGGACGAGCTGTAC  
 AAGTCCGGCCGGATCAGATCTCGAGCTCAAGCTTCAAGTTCGATCCAGATCTTAAGTAAGTAAAGCTCGAGAGCTGCTTTCTT  
 GCTGTCCAATTTCTATTAAAGGTTCCCTTTGTTCCCTAAGTCCAATCAACTAAGTCAACTGGGGGATATTATGAAGGGCCTTGAGCATC  
 TGGATTCTGCCTAATAAAAAACATTTATTTTTCATTGCTGCGTTCGAGAGCTCGCTTTCTTGTCTCCAATTTCTATTAAAGGTT  
 CCTTTGTTCCCTAAGTCCAACCTAAACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACAT  
 TTATTTTTCATTGCTGCGTTCGAGAGCTAA  
 AA-3'

**Figure S5.** Sequence of the eGFP-encoding mRNA transcript (5' to 3') used as a model system for long-read sequencing in this study. Structurally relevant sequence segments are colour coded in the following way: The ribosomal binding site in turquoise; the start codon in green; the stop codon in red; and the poly(A) tail in grey.

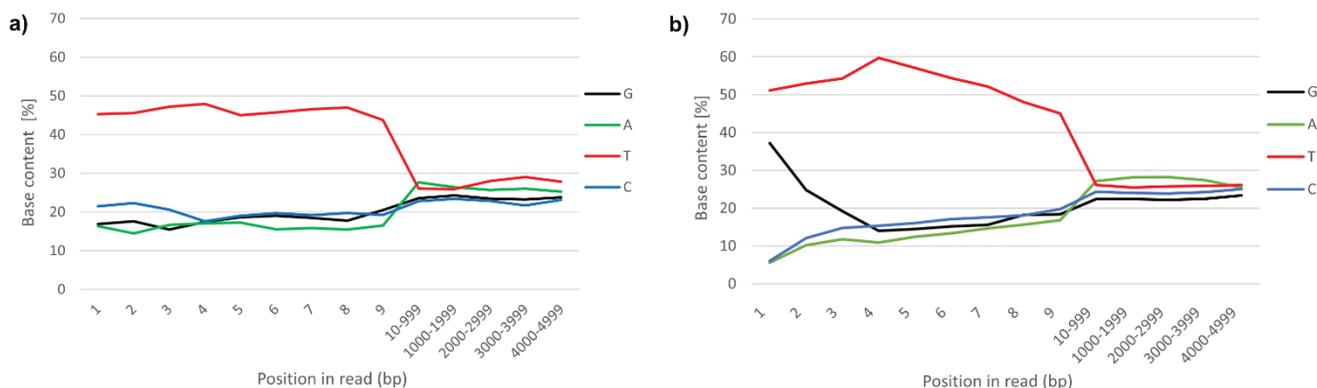
5' - **ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTTAG** CTCTCGACGCAGCAATGAAAATAAAT  
 GTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGACTTAGGGAAACAAAGGAA  
 CCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTCTCGACGCAGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGA  
 TGCTCAAGGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGACTTAGGGAAACAAAGGAACCTTTAATAGAAATTGGACAGCAA  
 GAAAGCGAGCTCTCGAGCTTACTTACTTAAGATCTGGATCAATTCGAAGCTTGAGCTCGAGATCTGAGTCCGGCCGGACTTGT  
 ACAGCTCGTCCATGCCGAGAGTGATCCCGGGCGGGTACGAACTCCAGCAGGACCATGTGATCGCGCTTCTCGTTGGGGTCT  
 TTGCTCAGGGCGGACTGGGTGCTCAGGTAGTGGTTGTGCGGCAGCAGCACGGGGCCGTCGCCGATGGGGGTGTTCTGCTGGTA  
 GTGGTCCGGCGAGCTGCACGCTGCCGTCCCTCGATGTTGTGGCGGATCTTGAAGTTCACCTTGATGCCGTTCTTCTGCTTGTTCGG  
 CCATGATATAGACGTTGTGGCTGTTGTAGTTGTACTCCAGCTTGTGCCCCAGGATGTTGCCGTCCCTCCTTGAAGTCGATGCC  
 TTCAGCTCGATGCCGTTACCAGGGTGTGCCCTCGAACTTCACCTCGGGCGGGTCTTGTAGTTGCCGTCGTCCCTTGAAGAA  
 GATGGTGCCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTGCTTTCATGTGGTCCGGGTAGCGGCTGA  
 AGCAGTCACGCCGTAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGT  
 GTCAGCTTGGCGTAGGTGGCATCGCCCTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGTTTACGTCGCCGTCAGCTCGAC  
 CAGGATGGGCACCACCCCGGTGAACAGCTCCTCGCCCTTGTCTACCATCTGCACGCCTCCTTGCTTACTAGTTCGCCC **CTAGA**  
**TCGGAAGAGCGTCTGTAGGGAAAGAGTGTAGATCTCCGGTGGTTCGCCGTATCATT** -3'

**Figure S6.** Sequence of the cDNA (5' to 3') obtained after reverse-transcription and 3'-adapter ligation of the eGFP-encoding mRNA transcript used as a model system for long-read sequencing in this study. Structurally relevant sequence segments are colour coded in the following way: The reverse transcription primer (RTP3) in turquoise; the added 3'-Azido-2',3'-ddG nucleoside in red; and the clicked adapter oligonucleotide (AA1) in yellow.

5' - **GACGCTCTTCCGATCTA** **C**GGGCGAACTAGTAAGCAAGGAGGCGTG...GAGAGNBAAAAAAAAAAAAAAAAAAGATCG **GA**  
**AGAGCGTCTGTAGGGAA** -3'

**Figure S7.** Sequence of the double stranded cDNA (5' to 3') obtained after PCR amplification of the click-ligated first-strand cDNA. Structurally relevant sequence segments are colour coded in the following way: The forward amplification primer (FP3) in yellow; the complementary C nucleotide of the added 3'-Azido-2',3'-ddG in red; and the reverse complementary sequence to the reverse amplification primer (RP3) in turquoise.

### Additional sequencing data analysis and enzyme information



**Figure S8.** Base-content at read-ends, generated with FastQC<sup>8</sup> from Jurkat ONT sequencing data. Individual percentages given for the first nine bases, averaged percentages for the remaining bins as indicated on the x-axis. Data was truncated after 4999 bases for clarity. **a)** Read data derived from click chemistry-based library preparation **b)** Read data derived from standard library preparation (TS).

Table S2. Enzymes and their respective suppliers, EC numbers, and research usage within in this study.

<b>Enzyme</b>	<b>Supplier</b>	<b>EC Number</b>	<b>Activity</b>
TranscriptAid Enzyme Mix	Thermo Fisher Scientific	EC 2.7.7.6	DNA directed RNA Polymerase
SuperScript <sup>™</sup> IV Reverse Transcriptase (200 U/ $\mu$ L)	Thermo Fisher Scientific	EC 2.7.7.49	RNA directed DNA Polymerase
RNaseOUT <sup>™</sup> Recombinant Ribonuclease Inhibitor	Thermo Fisher Scientific	not available	RNase Inhibition
RNase H (5 U/ $\mu$ L)	New England Biolabs	EC 3.1.26.4	Endoribonuclease
RNase A (10 mg/ $\mu$ L)	Thermo Fisher Scientific	EC 3.1.27.5	Endoribonuclease
Terminal deoxynucleotidyl Transferase TdT (20 U/ $\mu$ L)	Thermo Fisher Scientific	EC 2.7.7.31	Terminal Deoxynucleotidyl Transferase
LongAmp <sup>®</sup> Taq 2X Master Mix	New England Biolabs	EC 2.7.7.7	DNA Polymerase

## References

- (1) Elrod, N. D.; Jaworski, E. A.; Ji, P.; Wagner, E. J.; Routh, A. Development of Poly(A)-ClickSeq as a Tool Enabling Simultaneous Genome-Wide Poly(A)-Site Identification and Differential Expression Analysis. *Methods* **2019**, *155*, 20–29. <https://doi.org/10.1016/j.ymeth.2019.01.002>.
- (2) Picelli, S.; Faridani, O. R.; Björklund, Å. K.; Winberg, G.; Sagasser, S.; Sandberg, R. Full-Length RNA-Seq from Single Cells Using Smart-Seq2. *Nat. Protoc.* **2014**, *9* (1), 171–181. <https://doi.org/10.1038/nprot.2014.006>.
- (3) Han, Y.; Gao, X.; Liu, B.; Wan, J.; Zhang, X.; Qian, S. B. Ribosome Profiling Reveals Sequence-Independent Post-Initiation Pausing as a Signature of Translation. *Cell Res.* **2014**, *24* (7), 842–851. <https://doi.org/10.1038/cr.2014.74>.
- (4) Shepard, P. J.; Choi, E. A.; Lu, J.; Flanagan, L. A.; Hertel, K. J.; Shi, Y. Complex and Dynamic Landscape of RNA Polyadenylation Revealed by PAS-Seq. *RNA* **2011**, *17* (4), 761–772. <https://doi.org/10.1261/rna.2581711>.
- (5) Tang, F.; Barbacioru, C.; Wang, Y.; Nordman, E.; Lee, C.; Xu, N.; Wang, X.; Bodeau, J.; Tuch, B. B.; Siddiqui, A.; Lao, K.; Surani, M. A. MRNA-Seq Whole-Transcriptome Analysis of a Single Cell. *Nat. Methods* **2009**, *6* (5), 377–382. <https://doi.org/10.1038/nmeth.1315>.
- (6) Shi, X.; Jarvis, D. L. A New Rapid Amplification of cDNA Ends Method for Extremely Guanine plus Cytosine-Rich Genes. *Anal. Biochem.* **2006**, *356* (2), 222–228. <https://doi.org/10.1016/j.ab.2006.06.028>.
- (7) Croce, S.; Serdjukow, S.; Carell, T.; Frischmuth, T. Chemoenzymatic Preparation of Functional Click-Labeled Messenger RNA. *ChemBioChem* **2020**, *21* (11), 1641–1646. <https://doi.org/10.1002/cbic.201900718>.
- (8) Andrews, S. *FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]*. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed 2022-07-14).

# ChemBioChem

Supporting Information

## **Orthogonal End Labelling of Oligonucleotides through Dual Incorporation of Click-Reactive NTP Analogues**

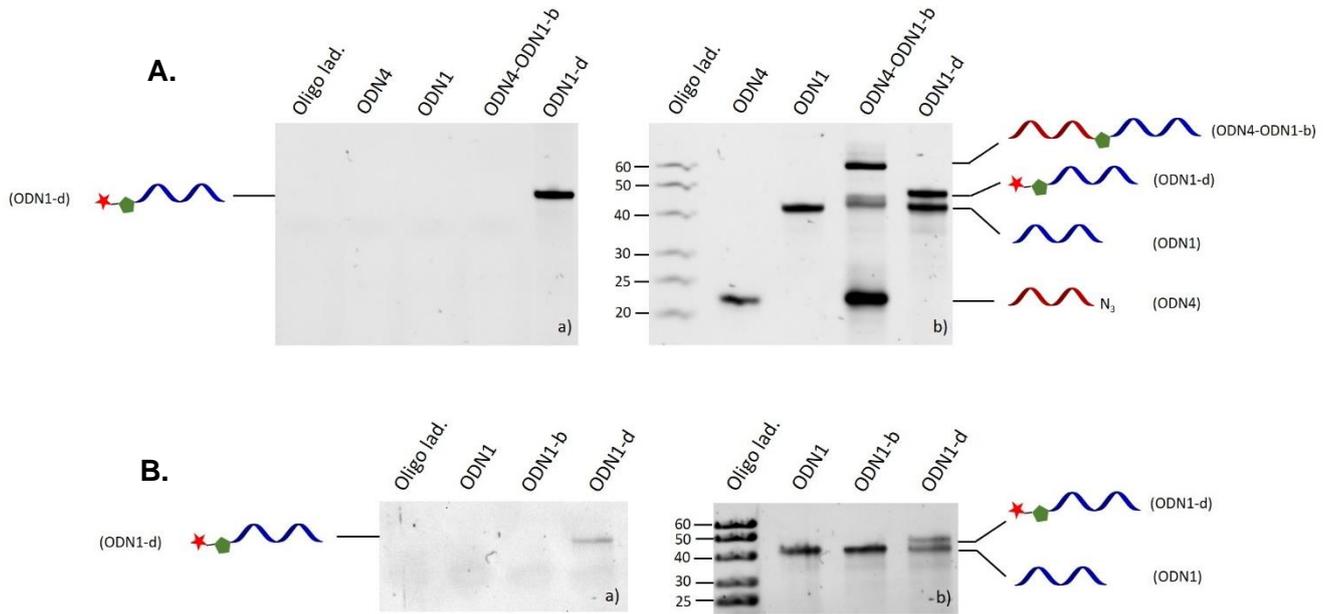
Eva S. Schönegger, Antony Crisp, Marco Radukic, Jonas Burmester, Thomas Frischmuth,\*  
and Thomas Carell\*

## Table of Contents

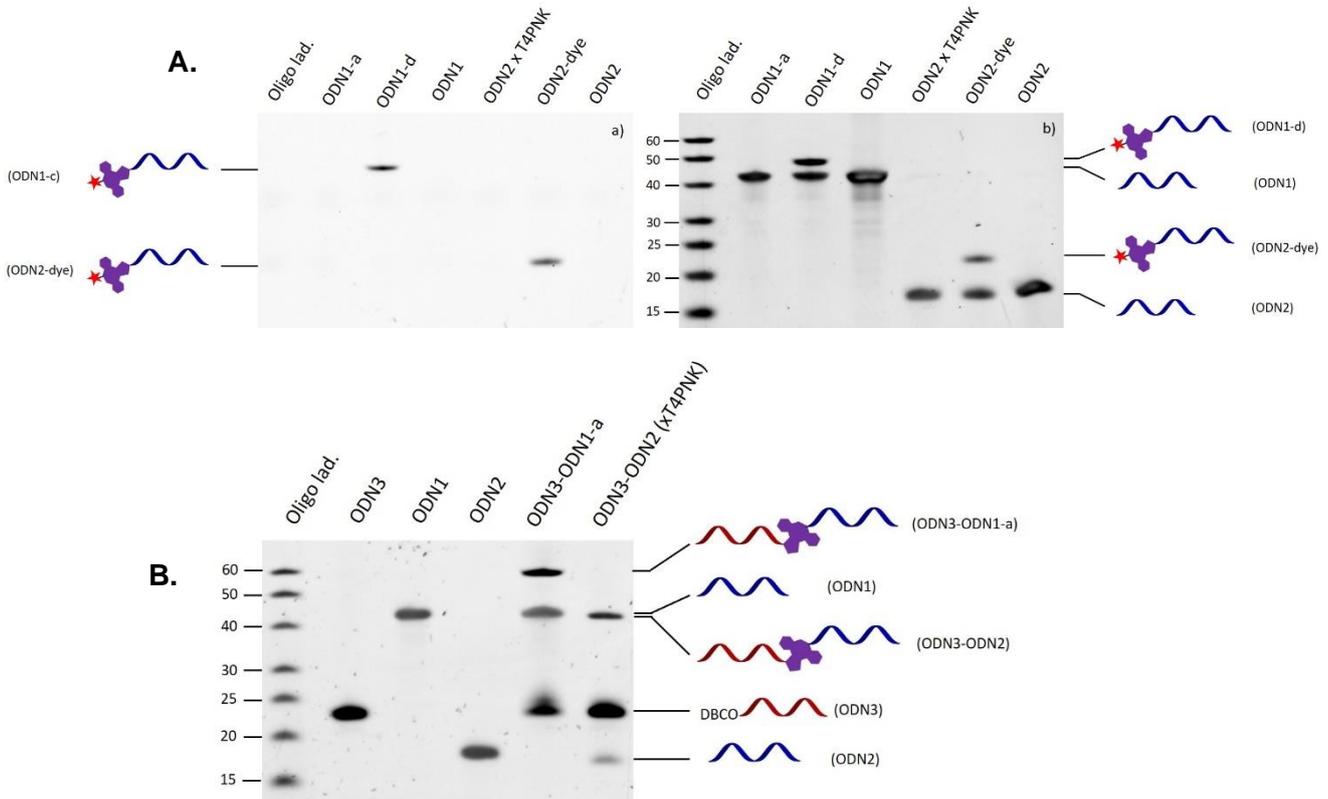
1. Supplementary Figures .....	2
2. Supplementary Tables.....	5
3. Supplementary Methods.....	10
3.1 General experimental methods .....	10
3.2 Synthesis and Characterization of $\gamma$ -phosphate modified ATP analogues .....	11
3.3 Synthesis and Characterization of unmodified DNA Oligonucleotides .....	17
3.4 T4 PNK catalysed 5'-end labelling reactions.....	18
3.4.1 Quantification of T4 PNK catalysed labelling reactions via analytical RP-HPLC.....	18
3.4.2 Impact of PEG on substrate conversion .....	18
3.4.3 Screening of $\gamma$ -phosphate modified ATP analogues.....	20
3.5 Click reactions with 5'-end labelled oligonucleotides .....	21
3.5.1 Cu(I) catalysed azide-alkyne cycloaddition (CuAAC) .....	21
3.5.2 Strain promoted azide-alkyne cycloaddition (SPAAC).....	22
3.5.3 Dual labelling and subsequent click reactions .....	23
3.5.4 Circularization of ODN5 via SPAAC.....	24
4. T4 PNK-mediated analogue hydrolysis assay (ADP-Glo Assay) .....	25
5. Molecular docking analysis.....	26
6. NMR spectra of synthesised compounds.....	28
7. References .....	39

# SUPPORTING INFORMATION

## 1. Supplementary Figures

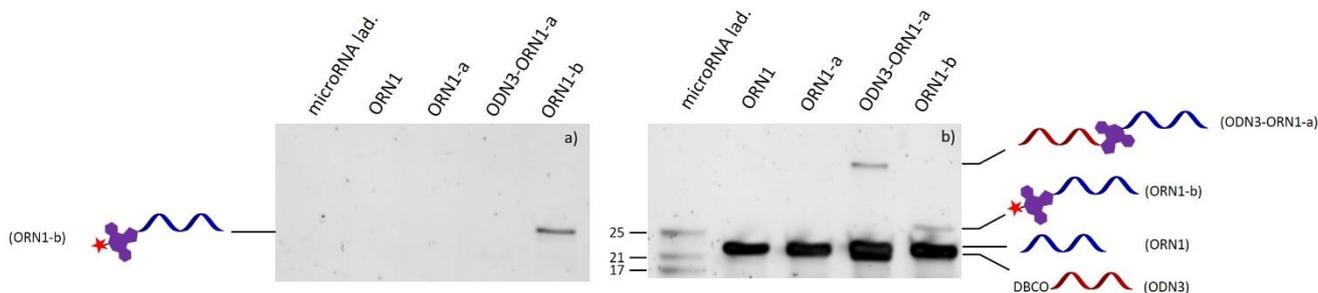


**Figure S1. A.** dPAGE (15%) analysis of ODN1-d and ODN4\_ODN1-b adduct (I). **a)** before staining **b)** after staining with GelRed® Nucleic Acid Stain. **B.** dPAGE (15%) analysis of ODN1-b and ODN1-d (II) **a)** before staining **b)** after staining with GelRed® Nucleic Acid Stain.

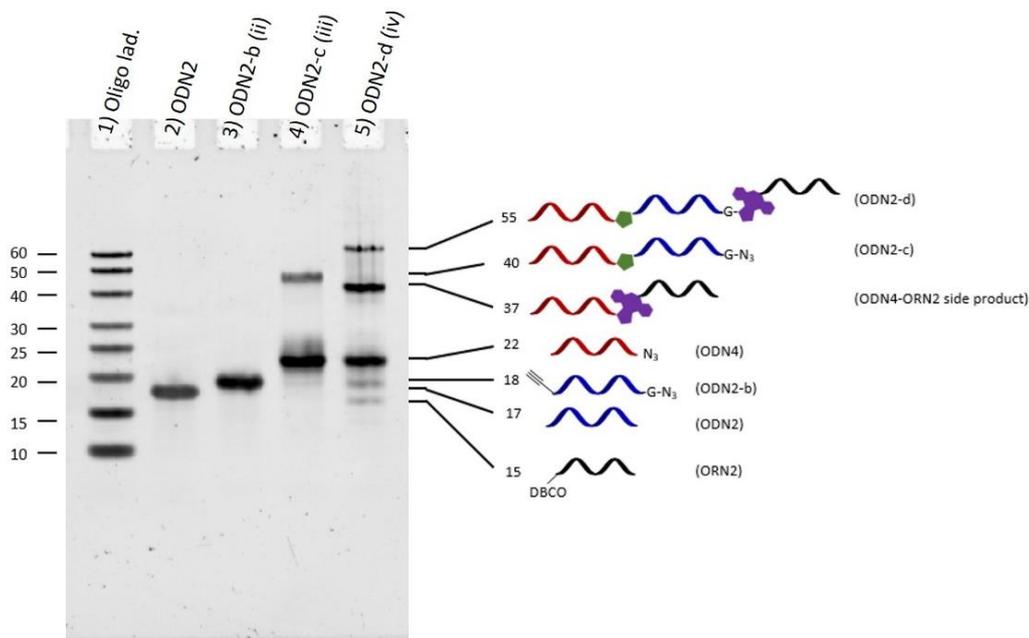


**Figure S2. A.** dPAGE (15%) analysis of T4 PNK reaction on ODN1 (ODN1-a) and ODN2 (ODN2 x T4 PNK), and analysis of subsequent SPAAC (ODN1-c and ODN2-dye). **a)** before staining **b)** after staining with GelRed® Nucleic Acid Stain. **B.** dPAGE (15%) analysis of ODN-ODN SPAAC after staining with GelRed® Nucleic Acid Stain.

## SUPPORTING INFORMATION



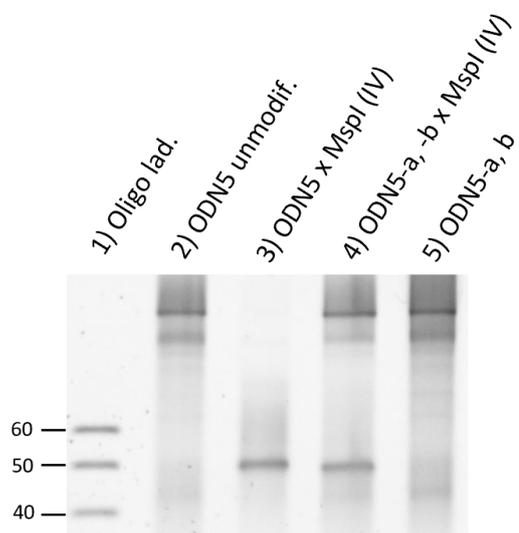
**Figure S3.** dPAGE (15%) analysis of T4 PNK reaction (ORN1-a) and subsequent SPAAC of ORN1 (ORN1-b and ODN3-ORN1-a). **a)** before staining **b)** after staining with GelRed® Nucleic Acid Stain.



**Figure S4.** dPAGE (15%) analysis of the individual steps from the dual labelling approach with subsequent click reactions. The gel image was taken after staining with GelRed® Nucleic Acid Stain. Lane 1) Oligo ladder. Lane 2) **ODN2** unmodified (17mer). Lane 3) **ODN2-b** (18mer). Lane 4) **ODN2-c** (40mer), excess of ODN4 (modified 22 mer) and unreacted ODN2-b. Lane 5) **ODN2-d** (55mer as DNA-DNA-RNA oligonucleotide assembly), side product (37mer, ODN4-ORN2, due to the excess of ODN4 and ODN2 (modified 15mer)) with excess of ODN4 (from step iii), unreacted ODN2-b (from step ii) and excess of ORN2 (from step vi). Detailed information of the individual steps is provided in section 3.5.3.

## SUPPORTING INFORMATION

---



**Figure S5.** 10% dPAGE of circularisation approach of ODN5. The gel image was taken after staining with GelRed® Nucleic Acid Stain. Lane 1) Oligo ladder. Lane 2) ODN5 (100mer) with a smaller MW- impurity from SPS. Lane 3) ODN5 digested with MspI as described in step VI. Lane 4) ODN5-a,b digested with MspI as described in step IV. Lane 5) ODN5-a,b (step III). Detailed information of the individual steps is provided in section 3.5.3.

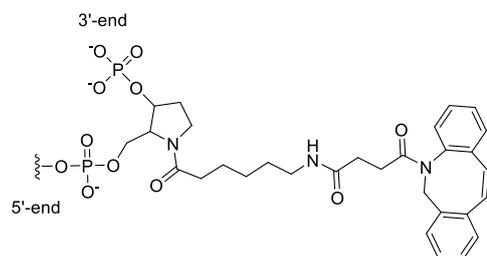
# SUPPORTING INFORMATION

## 2. Supplementary Tables

**Table S1:** Sequence of synthetic oligonucleotides. **ODN3, 4, 5, 6** and **ORN1-2** were purchased from Ella Biotech GmbH.

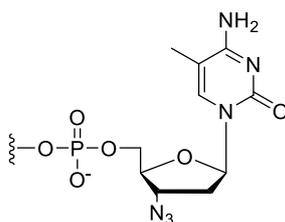
Name	Type	Sequence (5' to 3')	Modifications	Length (nt)
ODN1	DNA	GCC ATA GCG ATA GCG ATA CGC ACA CCC GGT CCT CCT AAT TTA T	-	43
ODN2	DNA	GGT CGC CGT ATC ATT AT	-	17
ODN3	modified DNA	TGG TAT CGC TAT CGC TAT GGC X	X = 3'-DBCO [a]	21
ODN4	modified DNA	TTG GTA TCG CTA TCG CTA TGG Z	Z = 3'-Azido-Methyl dC [b]	22
ODN5	DNA	GCT AGA TCG GAA GAG CGT CGG GCA GTA CAT CGA TCG TAT ATA CGC ACA CCC GGT CCT CCT GTC GCA TCA ATC GTA TCA TCC GAC GCT CTT CCG ATC TAG C	-	100
ODN6	DNA	AGG AGG ACC GGG TGT GCG TAT	-	21
ORN1	RNA	UUG GUA UCG CUA UCG CUA UGG	-	21
ORN2	modified RNA	YGU AUC GCU AUC GCU A	Y = 5'-DBCO [c] 3' = phosphate	15

[a] Structure of the modification X is given below.



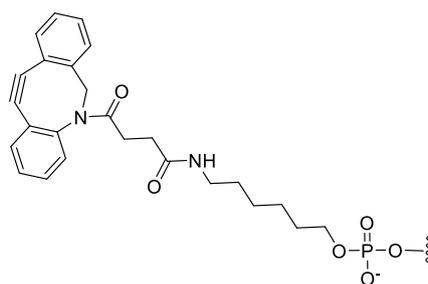
X = 3'-DBCO

[b] Structure of the modification Z is given below.



Z = 3'-Azido-Methyl dC

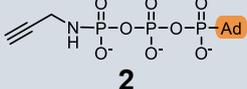
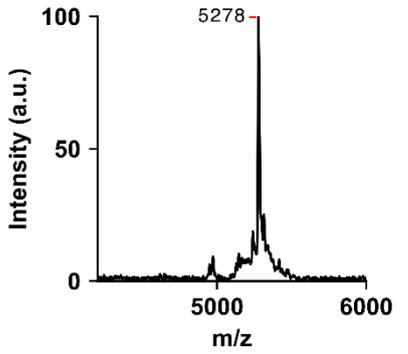
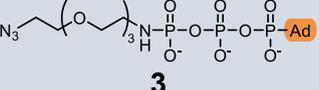
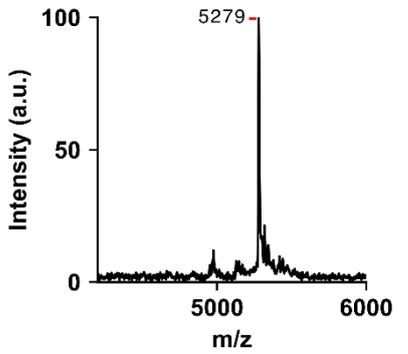
[c] Structure of the modification Y is given below.



Y = 5'-DBCO

## SUPPORTING INFORMATION

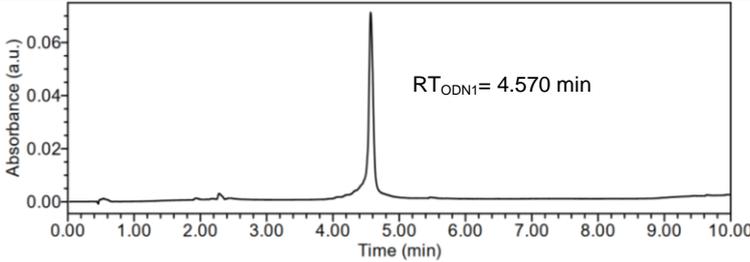
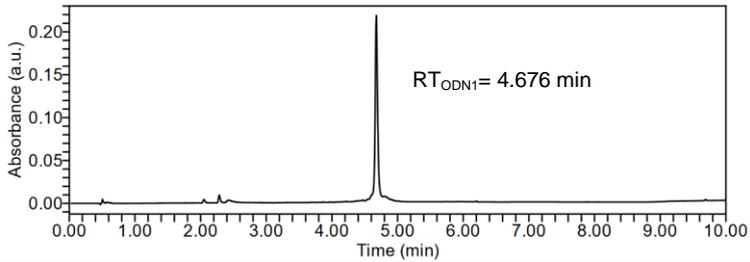
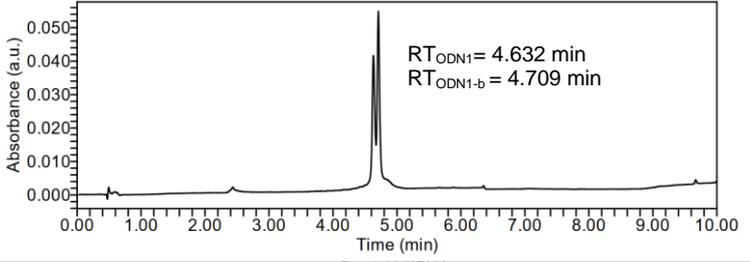
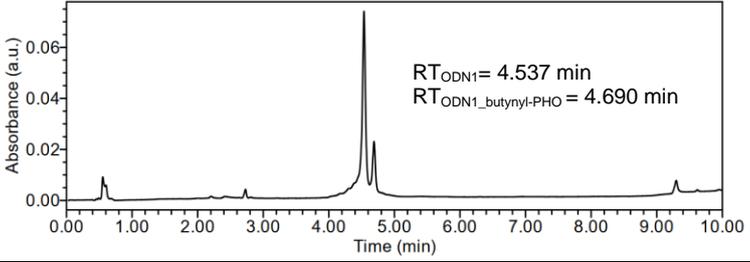
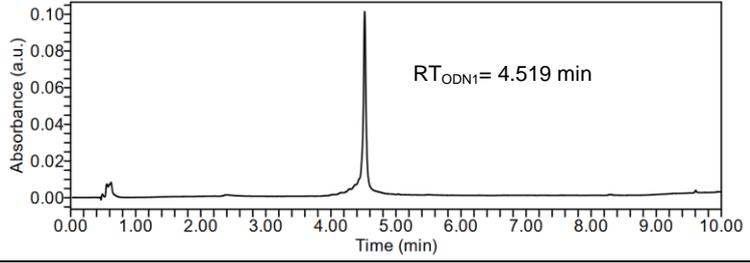
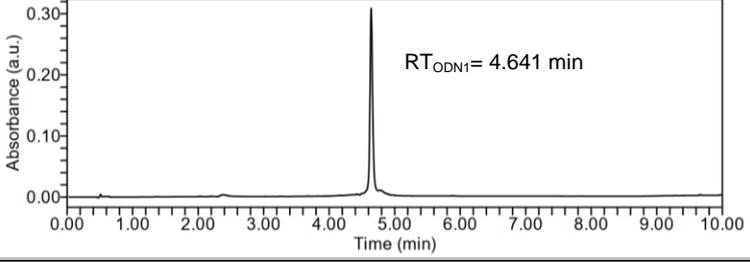
**Table S2.** MALDI-TOF analysis of ODN2 after T4 PNK catalyzed 5'-end labelling reaction with compound 2 and 3. The resulting spectra, indicating the  $m/z$  values of the phosphorylated product (5'-PHO-ODN2) only, suggest that the  $\gamma$ -modifications are not present in the ssDNA product.

 <p style="text-align: center;"><b>2</b></p>	<p><b>5'-PHO-ODN2</b> Sequence: 5'-PHO-GGT CGC CGT ATC ATT AT-3'</p> <p><math>[M_{5\text{'-PHO-2H+Na}}]^-</math> <math>m/z</math> calcd. 5277.36 found 5278.02</p>	
 <p style="text-align: center;"><b>3</b></p>	<p><b>5'-PHO-ODN2</b> Sequence: 5'-PHO-GGT CGC CGT ATC ATT AT-3'</p> <p><math>[M_{5\text{'-PHO-2H+Na}}]^-</math> <math>m/z</math> calcd. 5277.36 found 5279.77</p>	

Ad = 5'-adenosine

## SUPPORTING INFORMATION

**Table S3:** Conversion yield [%] of 5'-end labelling of ODN1 with  $\gamma$ -phosphate modified ATPs.  $RT_{ODN1\_N3\text{-ethyl-PHO=ODN1-a}}$ ,  $RT_{ODN1\_propargyl\text{-PHO=ODN1-b}}$

compound	Conversion yield of 5'-end labelling of ODN1 with $\gamma$ -phosphate modified ATP (compound 2-10) [%]	Analytical RP-HPL chromatogram (gradient: 0-30% buffer B)
<b>2</b>	0	 <p>RP-HPL chromatogram for compound 2. The x-axis is Time (min) from 0.00 to 10.00. The y-axis is Absorbance (a.u.) from 0.00 to 0.06. A single sharp peak is observed at <math>RT_{ODN1} = 4.570</math> min.</p>
<b>3</b>	0	 <p>RP-HPL chromatogram for compound 3. The x-axis is Time (min) from 0.00 to 10.00. The y-axis is Absorbance (a.u.) from 0.00 to 0.20. A single sharp peak is observed at <math>RT_{ODN1} = 4.676</math> min.</p>
<b>4</b>	56.65	 <p>RP-HPL chromatogram for compound 4. The x-axis is Time (min) from 0.00 to 10.00. The y-axis is Absorbance (a.u.) from 0.000 to 0.050. Two peaks are observed at <math>RT_{ODN1} = 4.632</math> min and <math>RT_{ODN1-b} = 4.709</math> min.</p>
<b>5</b>	19.25	 <p>RP-HPL chromatogram for compound 5. The x-axis is Time (min) from 0.00 to 10.00. The y-axis is Absorbance (a.u.) from 0.00 to 0.06. Two peaks are observed at <math>RT_{ODN1} = 4.537</math> min and <math>RT_{ODN1\_butynyl\text{-PHO}} = 4.690</math> min.</p>
<b>6</b>	0	 <p>RP-HPL chromatogram for compound 6. The x-axis is Time (min) from 0.00 to 10.00. The y-axis is Absorbance (a.u.) from 0.00 to 0.10. A single sharp peak is observed at <math>RT_{ODN1} = 4.519</math> min.</p>
<b>7</b>	0	 <p>RP-HPL chromatogram for compound 7. The x-axis is Time (min) from 0.00 to 10.00. The y-axis is Absorbance (a.u.) from 0.00 to 0.30. A single sharp peak is observed at <math>RT_{ODN1} = 4.641</math> min.</p>

## SUPPORTING INFORMATION

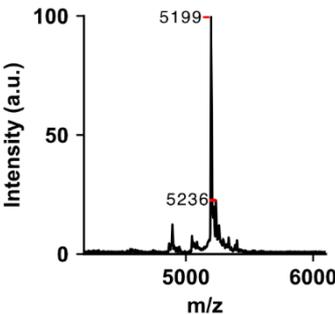
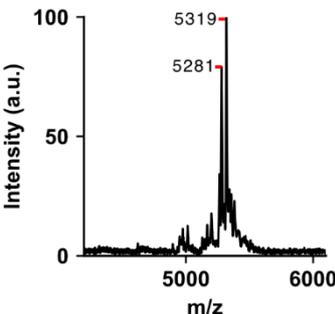
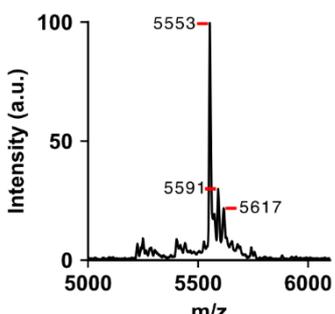
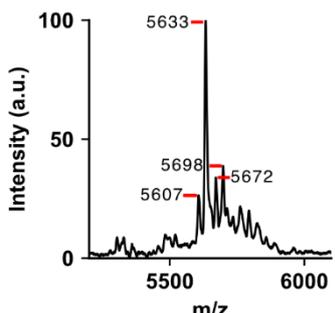
8	0	
9	36.82	
10	6.54	

**Table S4.** Analytical RP-HPL chromatograms of T4 PNK reaction with compound 9 on ODN1 (0% PEG) and subsequent SPAAC with DBCO-Sulfo-Cy3. . RT<sub>ODN1\_N3-ethyl-PHO</sub>=ODN1-a.

sample	Analytical RP-HPL chromatogram (gradient: 0-30% buffer B from 0 → 4 min, followed by 30-85%buffer B from 4 → 10 min)
unmodified ODN1	
T4 PNK with compound 9 on ODN1 with 0% PEG	
T4 PNK with compound 9 on ODN1 with 0% PEG and subsequent SPAAC with DBCO-Sulfo-Cy3 (OND1-c giving 2 regioisomers)	

## SUPPORTING INFORMATION

**Table S5.** MALDI-TOF analysis of ODN2, 3'-end and 5'-end labelled ODN2 respectively and dually labelled ODN2 (ODN2-b).

<p><b>ODN2 (unmodified)</b> Sequence: 5'-GGT CGC CGT ATC ATT AT-3'</p> <p>Fragments: [M-2H+Na]<sup>-</sup> m/z calcd. 5197.38 found 5198.54</p> <p>[M-3H+Na+K]<sup>-</sup> m/z calcd. 5235.47 found 5235.95</p>	
<p><b>ODN2_5'-propargyl-PHO</b> Sequence: 5'-<i>propargyl-PHO</i>-GGT CGC CGT ATC ATT AT-3'</p> <p>Fragments: [M-2H+Na]<sup>-</sup> m/z calcd. 5315.41 found 5318.87</p> <p>[M<sub>5'-PHO</sub>-2H+Na]<sup>-</sup> m/z calcd. 5277.36 found 5280.81</p>	
<p><b>ODN2-a</b> Sequence: 5'-GGT CGC CGT ATC ATT AT-G_N<sub>3</sub>-3'</p> <p>Fragments: [M-2H+Na]<sup>-</sup> m/z calcd. 5551.60 found 5553.29</p> <p>[M-3H+Na+K]<sup>-</sup> m/z calcd. 5589.69 found 5590.87</p>	
<p><b>ODN2-b and 5'-PHO-ODN2-a</b> Sequence: 5'-<i>propargyl-PHO</i>-GGT CGC CGT ATC ATT AT- G_N<sub>3</sub>-3'</p> <p>Fragments: [M-2H+Na]<sup>-</sup> m/z calcd. 5669.63 found 5671.49</p> <p>[M<sub>5'-PHO</sub>-2H+Na]<sup>-</sup> m/z calcd. 5631.58 found 5633.27</p>	

### 3. Supplementary Methods

#### 3.1 General experimental methods

**Materials and methods.** Chemicals were purchased from Sigma-Aldrich/Merck, BroadPharm, ABCR, Acros Organics or VWR and were used without further purification. Solvents were of reagent grade or purchased in septum-sealed bottles stored under an inert atmosphere. Unless otherwise specified, all reactions were magnetically stirred under a positive pressure of Argon (Ar). Reactions and chromatography fractions were monitored by qualitative thin-layer chromatography (TLC) on silica gel F254 TLC plates from Merck KGaA and visualised by UV illumination or monitored by analytical RP-HPLC (see analytical and preparative RP HPLC). Dialysis membrane Spectra/Por® Biotech CE MWCO 100–500 Da (10 mm surface width) was purchased from Repligen and used as received.

**Nuclear magnetic resonance (NMR).** NMR spectra were recorded on a Bruker Avance III HD 400 (400 MHz), Bruker Avance Neo (500 MHz) or Bruker Avance III (800 MHz) spectrometers. <sup>1</sup>H NMR shifts were calibrated to their residual protons of the deuterated solvent: CD<sub>2</sub>HOD (3.31 ppm), D<sub>2</sub>O (4.79 ppm). <sup>13</sup>C NMR shifts were calibrated to the residual solvent: CD<sub>3</sub>OD (49.00 ppm). The chemical shifts ( $\delta$ ) are given in ppm, the coupling constants (J) in Hz. Multiples are abbreviated as follows: s= singlet, d= doublet, t= triplet, q= quartet, dd= doublet of doublets, m= multiplet. The numbering in the assignments does not follow IUPAC rules and neither residual solvent signals nor tris-triethylammonium salt signals (TEAA buffer from RP-HPLC purification: quartet around 2.90-3.20 ppm, singlet around 1.90 ppm and triplet around 1.20-1.30 ppm) have been assigned for clarity. All NMR spectra were analysed using the software MestReNova 14.1.1 from Mestrelab Research S. L.

**Mass Spectrometry (MS).** High resolution mass spectra (HRMS (ESI)) were recorded by the analytical section of the Department of Chemistry of the Ludwig-Maximilians-Universität München on a MAT 90 (ESI) spectrometer from Thermo Finnigan GmbH. Matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectra were recorded on an autoflex® maX. Before MALDI-TOF MS, the samples (2  $\mu$ L) were desalted on a 0.025  $\mu$ M VSWP filter (Millipore) and 1  $\mu$ L of the desalted sample was applied to a stainless-steel sample plate using dried droplet method with a 3-hydroxypicolinic acid matrix (1  $\mu$ L, HPA: 25 mg 3-hydroxypicolinic acid, 5 mg ammonium citrate, 5  $\mu$ L 15-crown-5 in 0.5 mL H<sub>2</sub>O/MeCN= 1:1). The obtained spectra were analysed with FlexAnalysis 3.4 from Bruker Daltonics and processed with Affinity Designer.

**Reversed-Phase High-Performance Liquid Chromatography (RP-HPLC).** Semi-preparative RP-HPLC was performed on a 1260 Infinity II Manual Preparative LC System from Agilent (Agilent Infinity II 12260 VWD: G7114A) equipped with the column VP 250/10 Nucleodur 100-5 C18 ec from Macherey-Nagel. A flow rate of 5 mL/min with varying gradients between 0-10% and 0-60% of buffer B over 45 min was applied for purification. The following buffer system was used: buffer A: 100 mM NEt<sub>3</sub>/HOAc (pH 7.0) in H<sub>2</sub>O and buffer B: 100 mM in NEt<sub>3</sub>/HOAc (pH 7.0) 80% (v/v) acetonitrile. Analytical RP-HPLC was performed on a Waters Alliance (2695 Separation Module, 2998 Photodiode Array Detector) equipped with a XBridge® Oligonucleotide BEH C18 Column (30 Å, 2.5  $\mu$ m, 2.1 mm x 50 mm) using a flow rate of 1.5 mL and the same buffer system as mentioned above. Unless otherwise specified, varying gradients, between 0-10% and 0-60% buffer B from 0  $\rightarrow$  8 min, followed by 10-85% up to 60-85% buffer B from 8  $\rightarrow$  10 min, were applied.

**Determination of ATP and oligonucleotide concentrations.** The absorbance of ATP tetrabutylammonium salt,  $\gamma$ -phosphate modified ATP analogues and synthesised oligonucleotides was measured at 260 nm using an Implen NanoPhotometer® P300. The extinction coefficients ( $\epsilon$ ) of the oligonucleotides were calculated using the OligoAnalyzer™ Tool from Integrated DNA Technologies. For ATP tetrabutylammonium salt and for compounds 4 - 8 an assumed extinction coefficient of  $\epsilon = 15400$  L mol<sup>-1</sup> cm<sup>-1</sup> was applied, as these modifications are not expected to substantially alter the absorption

## SUPPORTING INFORMATION

properties of unmodified ATP. The concentrations of the oligonucleotides and ATP were further calculated *via* Beer-Lambert law.

**Denaturing Polyacrylamide Gel Electrophoresis (dPAGE) Analysis.** Unmodified RP-HPLC purified oligonucleotides and spin column purified oligonucleotide products after T4 PNK-catalysed labelling (section 5) were separated by either 10 or 15% denaturing urea (7 M) polyacrylamide gel electrophoresis in Tris-Borate-EDTA (TBE) buffer (1X: 0.13 mM tris, pH 7.6, 45 mM boric acid, 2.5 mM EDTA) at constant voltage (150 or 200 V) for 40 min - 1.5 h. The gels were stained for 30 min using GelRed® Nucleic Acid Stain (Millipore) according to the manufacturers protocol. The bands were detected on Gel Doc™ EZ Imager (Bio-Rad) using a UV Sample Tray (#1708271MP from Bio-Rad) and analysed with Image Lab™ software. Single-stranded DNA oligo length standard (10/60 ladder) was purchased from Integrated DNA Technologies, microRNA marker was purchased from New England Biolabs. RNA loading dye (2X) was purchased from Thermo Fisher Scientific.

### 3.2 Synthesis and Characterization of $\gamma$ -phosphate modified ATP analogues

#### General procedure 1

Conversion of ATP disodium salt into ATP tetrabutylammonium salt. For the synthesis of  $\gamma$ -O-linked ATP analogues *via* alkylation of the  $\gamma$ -phosphate, reported by Hacker *et al.* the ATP disodium salt needs to be converted into its tetrabutylammonium salt to guarantee strictly anhydrous conditions.<sup>[1]</sup> Therefore, 4 g Chelex® 100 sodium form was resuspended in 20 mL H<sub>2</sub>O, transferred into a glass column and washed with 500 mL H<sub>2</sub>O. Then 10 g tetrabutylammonium bromide was dissolved in 100 mL H<sub>2</sub>O, flushed through the column and washed with 500 mL H<sub>2</sub>O. Subsequently, 300-550  $\mu$ mol of ATP disodium salt were diluted in 10 mL H<sub>2</sub>O, and applied to the column. The column was washed with ca. 90mL H<sub>2</sub>O and fractions of 5 mL were collected and identified by spotting on TLC plates and visualization *via* UV lamp. The fractions containing the nucleotide were combined and evaporated under reduced pressure. The residue was dissolved in 10 mL H<sub>2</sub>O and the absorbance of an aliquot of a 1:10 dilution at 260 nm was measured. Until further usage the nucleotide solution was stored at -20°C.

#### General procedure 2

Syntheses of  $\gamma$ -O linked ATP analogues were performed in reference to Hacker *et al.* with adjustments.<sup>[1-2]</sup> Therefore, 100-300  $\mu$ mol of ATP Bu<sub>4</sub>N<sup>+</sup> salt was evaporated under reduced pressure, dissolved in 5 mL anhydrous DMF and evaporated under reduced pressure. This step was repeated two times and the residue was dried on high vacuum overnight. Further, 2.7 - 6.0 equiv. of alkylating agent was dissolved in 2-6 mL anhydrous DMF, 1-2 g oven-dried 4 Å molecular sieves were added and stirred at room temperature overnight under Argon atmosphere. The next day the dried nucleotide was dissolved in 2-6 mL anhydrous DMF, flushed with Argon and equipped with a septum and Argon balloon in order to guarantee anhydrous conditions. Subsequently, the alkylating agent in DMF was added dropwise, under rigorous stirring, to the ATP solution and stirred overnight at room temperature (r.t.). Reaction control was performed *via* analytical RP-HPLC. Unless further conversion was observed, the solvent was evaporated under reduced pressure. The residue was dissolved in 10 mL HPLC grade H<sub>2</sub>O and centrifuged (Hermle Labortechnik Centrifuge Z 206A) at 6000 rpm for 10 min at r.t. The supernatant was purified two to three times *via* semi-preparative RP-HPLC. The product fractions were identified by ESI-MS and freeze dried. Since T4 PNK is inhibited by ammonium ions<sup>[3],[4]</sup> the  $\gamma$ -O linked ATP triethylammonium salts were, unless further specified, transferred into their sodium salt by precipitation (General procedure 4). The conversion yield [%] of the isolated ATP analogues was determined by analytical RP-HPLC.

#### General procedure 3

Syntheses of  $\gamma$ -N linked ATP analogues were performed in reference to Serdjukov *et al.* with adjustments.<sup>[5]</sup> Therefore, in a 1.5 mL reaction tube, ATP disodium salt hydrate (20.0 mg, 36.29  $\mu$ mol, 1.0 equiv.) was dissolved in H<sub>2</sub>O (170-363  $\mu$ L), followed by the addition of EDC·HCl (34.8 mg, 181.4

## SUPPORTING INFORMATION

---

$\mu\text{mol}$ , 5.0 eq.). The reaction was mixed on a thermomixer at 25°C for 30 min (600 rpm). Then, the amine (5.0 equiv.) dissolved either in  $\text{H}_2\text{O}$ , or in DMF, was added dropwise to the reaction mixture and mixed for further 18 h (25°C, 600 rpm). Reaction control was performed *via* analytical RP-HPLC. The solvent was evaporated under reduced pressure and the residue was purified *via* semi-preparative RP-HPLC to yield the tris-triethylammonium salts of the  $\gamma$ -nitrogen linked ATP analogues. The conversion yield [%] of the isolated ATP analogues was determined by analytical RP-HPLC.

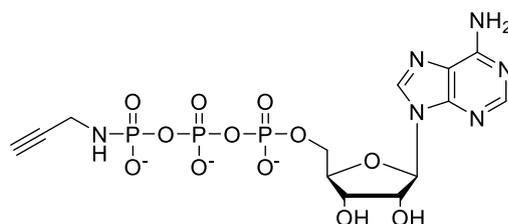
For the 5'-end labelling reactions the nucleotide tris-triethylammonium salts were converted into their sodium salts by ion exchange (Chelex® 100 sodium form), followed by 12 h of dialysis.

### General procedure 4

Conversion of  $\gamma$ -modified ATP tris-triethylammonium salt to sodium salt was either performed by ion exchange (Chelex® 100 sodium form), followed by 12 h of dialysis, or by precipitation. For the precipitation the lyophilized ATP tris-triethylammonium salt (ca. 2 mg) was dissolved in 200  $\mu\text{L}$  HPLC grade MeOH and added dropwise solution of anhydrous  $\text{NaClO}_4$  in HPLC grade acetone (32 mg in 1.8 mL). A colourless solid precipitated after the mixture was stored in the -20°C fridge for ca. 30 min. The supernatant was removed and the pellet was washed three times with 1 mL HPLC grade acetone. The final colourless solid was dried in high vacuum.

## SUPPORTING INFORMATION

### Synthesis of $\gamma$ -propargylamino-ATP (**2**)



The synthesis of compound **2** was performed with propargylamine (9.9 mg, 181.4  $\mu$ mol, 5.0 equiv.) dissolved in 57  $\mu$ L DMF as describes in general procedure 3. RP-HPLC purification afforded the tris-triethylammonium salt of **2** as a colourless oil after lyophilization (97%).

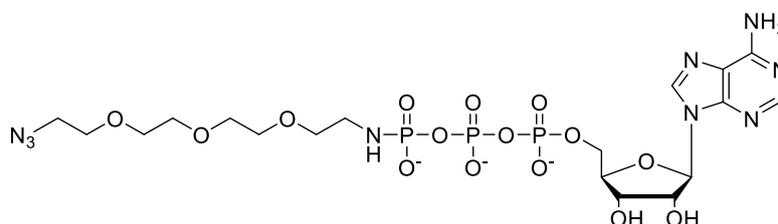
**$^1\text{H}$  NMR** (400 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  8.51 (s, 1H, H2), 8.23 (s, 1H, H8), 6.12 (d,  $J$  = 6.1 Hz, 1H, H1'), 4.77 – 4.75 (m, 1H, overlapping with HOD, H2'), 4.59 – 4.51 (m, 1H, H3'), 4.40 – 4.33 (m, 1H, H4'), 4.25 – 4.14 (m, 2H, H5'), 3.62 (d,  $J$  = 10.0, 2.4 Hz, 2H,  $\text{CH}_2$ ), 2.42 (t,  $J$  = 2.5 Hz, 1H, CH).  **$^{13}\text{C}$  NMR { $^1\text{H}$ }** (101 MHz,  $\text{D}_2\text{O}$ )  $\delta$  155.58, 152.76, 149.13, 139.80, 119.92, 86.56, 84.04, 82.66, 74.17, 71.10, 70.32, 65.10, 27.84.  **$^{31}\text{P}$  { $^1\text{H}$ }** NMR (162 MHz,  $\text{D}_2\text{O}$ )  $\delta$  -2.93 (d,  $J$  = 20.7 Hz, 1P,  $\gamma$ ), -11.56 (d,  $J$  = 20.3 Hz, 1P,  $\alpha$ ), -23.08 (t,  $J$  = 20.1 Hz, 1P,  $\beta$ ).

**HRMS (ESI):** calculated for  $\text{C}_{13}\text{H}_{18}\text{N}_6\text{O}_{12}\text{P}_3^-$  [ $\text{M}-\text{H}$ ] $^-$ : 543.0201; found: 543.0203

**Analytical RP-HPLC** (gradient: 0-5% Buffer B\*)  $R_t$  = 6.792 min

\* 0-5% Buffer B from 0  $\rightarrow$  10 min, followed by 5-25% from 10  $\rightarrow$  10.5 min

### Synthesis of $\gamma$ -11-azido-2,6,9-trioxaundecan-1-amino-ATP (**3**)



The synthesis of compound **3** was performed with 2-[2-[2-(2-Azidoethoxy)ethoxy]ethoxy]ethylamine (39.6 mg, 181.4  $\mu$ mol, 5.0 eq.) dissolved in 170  $\mu$ L DMF as describes in general procedure 3. RP-HPLC purification afforded the tris-triethylammonium salt of **3** as a colourless oil after lyophilization (48%).

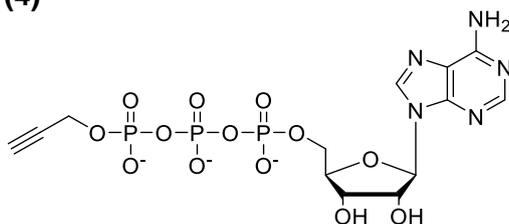
**$^1\text{H}$  NMR** (400 MHz,  $\text{D}_2\text{O}$ )  $\delta$  8.53 (s, 1H, H2), 8.25 (s, 1H, H8), 6.12 (d,  $J$  = 6.2 Hz, 1H, H1'), 4.77 – 4.75 (m, 1H, overlapping with HOD, H2'), 4.63 – 4.49 (m, 1H, H3'), 4.45 – 4.31 (m, 1H, H4'), 4.33 – 4.12 (m, 2H, H5'), 3.76 – 3.44 (m, 16H, 8x $\text{CH}_2$ ).  **$^{13}\text{C}$  NMR { $^1\text{H}$ }** and  **$^{31}\text{P}$  NMR { $^1\text{H}$ }** according to Lee *et al.*<sup>[6]</sup>

**HRMS (ESI):** calculated for  $\text{C}_{18}\text{H}_{31}\text{N}_9\text{O}_{15}\text{P}_3^-$  [ $\text{M}-\text{H}$ ] $^-$ : 706.1158; found: 706.11587

**Analytical RP-HPLC** (gradient: 0-30% buffer B)  $R_t$  = 4.205 min

## SUPPORTING INFORMATION

### Synthesis of $\gamma$ -propargyl-ATP (4)



The synthesis of compound **4** was performed with ATP tetrabutylammonium salt (123.2 mg, 100  $\mu$ mol, 1 equiv.), and propargyl p-toluenesulfonate (110.7 mg, 500  $\mu$ mol, 5.0 equiv.) as described in general procedure 1 and 2. RP-HPLC purification afforded the tris-triethylammonium salt of **4** which was further transferred to its sodium salt (general procedure 4) and isolated as a colourless solid after lyophilization (64%).

**$^1\text{H}$  NMR** (500 MHz, MeOD)  $\delta$  8.64 (s, 1H, H2), 8.19 (s, 1H, H8), 6.11 (d,  $J$  = 5.6 Hz, 1H, H1'), 4.72 – 4.63 (m, 3H, H2', CH<sub>2</sub>), 4.58 – 4.53 (m, 1H, H3'), 4.31 – 4.17 (m, 3H, H4', H5'), 2.75 (t,  $J$  = 2.5 Hz, 1H, CH).

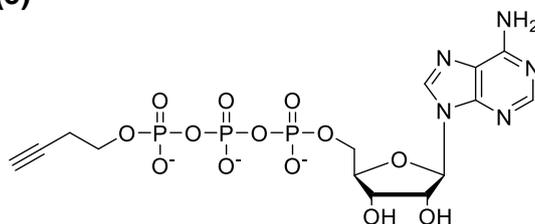
**$^{13}\text{C}$  { $^1\text{H}$ } NMR** (201 MHz, MeOD)  $\delta$  152.33, 139.54, 87.47, 84.23, 75.04, 70.48, 64.91, 53.37 (measured via HSQC experiment)

**$^{31}\text{P}$  { $^1\text{H}$ } NMR** (202 MHz, MeOD)  $\delta$  -10.14 (d,  $J$  = 17.5 Hz, 1P,  $\gamma$ ), -10.33 (d,  $J$  = 17.5 Hz, 1P,  $\alpha$ ), -20.40 (t,  $J$  = 17.5 Hz, 1P,  $\beta$ ).

**HRMS (ESI):** calculated for C<sub>13</sub>H<sub>15</sub>N<sub>5</sub>Na<sub>2</sub>O<sub>13</sub>P<sub>3</sub><sup>-</sup> [M-3H+2Na]<sup>-</sup>: 587.9680; found: 587.9681

**Analytical RP-HPLC** (gradient: 0-10% buffer B) R<sub>t</sub> = 3.658 min

### Synthesis of $\gamma$ -butynyl-ATP (5)



The synthesis of compound **5** was performed with ATP tetrabutylammonium salt (369.5 mg, 300  $\mu$ mol, 1 equiv.), and 3-butynyl p-toluenesulfonate (403.7 mg, 1.80 mmol, 6 equiv.) as described in general procedure 1 and 2. RP-HPLC purification afforded the tris-triethylammonium salt of **5** which was further transferred to its sodium salt (general procedure 4) and isolated as a colourless solid after lyophilization (15%).

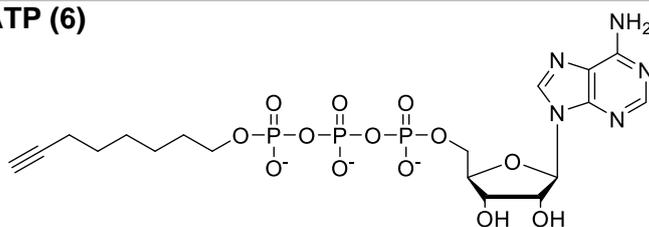
**$^1\text{H}$  NMR** (500 MHz, MeOD)  $\delta$  8.62 (s, 1H, H2), 8.17 (s, 1H, H8), 6.10 (d,  $J$  = 5.1 Hz, 1H, H1'), 4.67 – 4.61 (m, 1H, H2'), 4.55 – 4.49 (m, 1H, H3'), 4.29 – 4.20 (m, 3H, H4', H5'), 4.07 – 3.99 (m, 2H, CH<sub>2</sub>-O-P), 2.56 – 2.48 (m, 2H, CH<sub>2</sub>), 2.20 (t,  $J$  = 2.7 Hz, 1H, CH).  **$^{13}\text{C}$  { $^1\text{H}$ } NMR** (201 MHz, MeOD)  $\delta$  157.29, 153.71, 150.71, 141.60, 120.18, 89.19, 81.49, 76.51, 71.81, 70.70, 66.30, 65.28, 21.39, 21.35.  **$^{31}\text{P}$  { $^1\text{H}$ } NMR** (202 MHz, MeOD)  $\delta$  -9.60 (d,  $J$  = 16.1 Hz, 1P,  $\gamma$ ), -9.72 (d,  $J$  = 16.1 Hz, 1P,  $\alpha$ ), -19.50 (t,  $J$  = 16.0 Hz, 1P,  $\beta$ ).

**HRMS (ESI):** calculated for C<sub>14</sub>H<sub>17</sub>N<sub>5</sub>Na<sub>2</sub>O<sub>13</sub>P<sub>3</sub><sup>-</sup> [M-3H+2Na]<sup>-</sup>: 601.9837; found: 601.9835

**Analytical RP-HPLC** (gradient: 0-10% buffer B) R<sub>t</sub> = 4.681 min

## SUPPORTING INFORMATION

### Synthesis of $\gamma$ -octynyl-ATP (**6**)



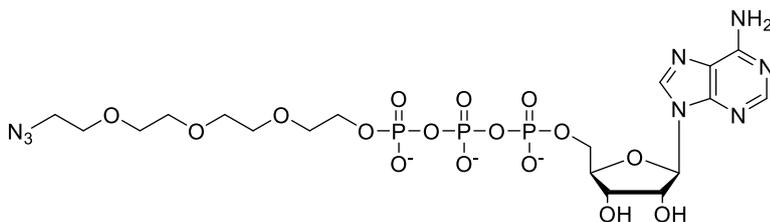
The synthesis of compound **6** was performed with ATP tetrabutylammonium salt (123.2 mg, 100  $\mu$ mol, 1 equiv.), and 8-bromooct-1-yne (60.1 mg, 318  $\mu$ mol, 3.18 equiv.) as described in general procedure 1 and 2. RP-HPLC purification afforded the tris-triethylammonium salt of **6** which was further transferred to its sodium salt (general procedure 4) and isolated as a colourless solid after lyophilization (68%).

**$^1\text{H}$  NMR** (500 MHz, MeOD)  $\delta$  8.64 (s, 1H, H2), 8.16 (s, 1H, H8), 6.09 (d,  $J$  = 5.0 Hz, 1H, H1'), 4.70 – 4.57 (m, 1H, H2'), 4.55 – 4.50 (m, 1H, H3'), 4.29 – 4.21 (m, 3H, H4', H5'), 3.95 – 3.87 (m, 2H, CH<sub>2</sub>-O-P), 2.18 – 2.08 (m, 3H, CH<sub>2</sub>, CH), 1.59 – 1.52 (m, 2H, CH<sub>2</sub>), 1.44 – 1.28 (m, 6H, 3xCH<sub>2</sub>).  **$^{13}\text{C}$  { $^1\text{H}$ } NMR** (201 MHz, MeOD)  $\delta$  157.29, 153.71, 150.69, 141.67, 120.16, 89.20, 85.48, 85.06, 76.57, 71.81, 69.36, 67.19, 66.27, 31.59, 29.69, 29.56, 26.32, 18.95.  **$^{31}\text{P}$  { $^1\text{H}$ } NMR** (202 MHz, MeOD)  $\delta$  -8.92 (d,  $J$  = 15.9 Hz, 1P,  $\gamma$ ), -9.65 (d,  $J$  = 15.9 Hz, 1P,  $\alpha$ ), -19.30 (t,  $J$  = 16.0 Hz, 1P,  $\beta$ ).

**HRMS (ESI):** calculated for C<sub>18</sub>H<sub>25</sub>N<sub>5</sub>Na<sub>2</sub>O<sub>13</sub>P<sub>3</sub><sup>-</sup> [M-3H+2Na]<sup>-</sup>: 658.0463; found: 658.0465

**Analytical RP-HPLC** (gradient: 0-30% buffer B) R<sub>t</sub> = 5.492 min

### Synthesis of $\gamma$ -(2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)-ATP (**7**)



The synthesis of compound **7** was performed with ATP tetrabutylammonium salt (123.2 mg, 100  $\mu$ mol, 1 equiv.), and 1-azido-2-[2-[2-(2-bromoethoxy)ethoxy]ethoxy]ethane (75.0 mg, 265.8  $\mu$ mol, 2.66 equiv.) as described in general procedure 1 and 2. RP-HPLC purification afforded the tris-triethylammonium salt of **7** which was further transferred to its sodium salt (general procedure 4) and isolated as a colourless solid after lyophilization (46%).

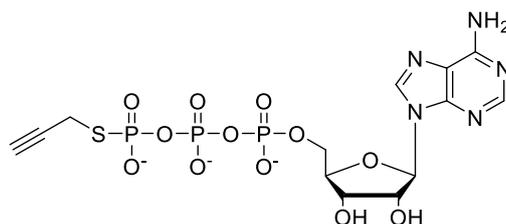
**$^1\text{H}$  NMR** (500 MHz, D<sub>2</sub>O)  $\delta$  8.50 (s, 1H, H2), 8.22 (s, 1H, H8), 6.10 (d,  $J$  = 6.2 Hz, 1H, H1'), 4.74 (dd,  $J$  = 6.1, 5.2 Hz, 1H, H2'), 4.54 – 4.51 (m, 1H, H3'), 4.38 – 4.31 (m, 1H, H4'), 4.23 – 4.16 (m, 2H, H5'), 4.05 – 3.99 (m, 2H, CH<sub>2</sub>-O-P), 3.66 – 3.61 (m, 4H, 2xCH<sub>2</sub>), 3.61 – 3.56 (m, 8H, 4xCH<sub>2</sub>), 3.43 (t,  $J$  = 4.7 Hz, 2H, CH<sub>2</sub>-N<sub>3</sub>).  **$^{31}\text{P}$  { $^1\text{H}$ } NMR (202 MHz, D<sub>2</sub>O)**  $\delta$  -11.13 (d,  $J$  = 19.0 Hz, 1P,  $\gamma$ ), -11.48 (d,  $J$  = 19.0 Hz, 1P,  $\alpha$ ), -23.21 (t,  $J$  = 19.2 Hz, 1P,  $\beta$ ).

**HRMS (ESI):** calculated for C<sub>18</sub>H<sub>28</sub>N<sub>8</sub>Na<sub>2</sub>O<sub>16</sub>P<sub>3</sub><sup>-</sup> [M-3H+2Na]<sup>-</sup>: 751.0637; found: 751.0631

**Analytical RP-HPLC** (gradient: 0-30% buffer B) R<sub>t</sub> = 4.340 min

## SUPPORTING INFORMATION

### Synthesis of $\gamma$ -[(propargyl)-thio]-ATP (**8**)



The synthesis of compound **8** was performed in reference to Espinasse *et al.* with adjustments.<sup>[7]</sup> Therefore, ATPyS tetra lithium salt (0.72 mg, 1.32  $\mu$ mol, 1.0 equiv.) was dissolved in 1.5 mL D<sub>2</sub>O and 25  $\mu$ L anhydrous DMF and stirred for 1 min. Then NaHCO<sub>3</sub> (0.70 mg, 8.33  $\mu$ mol, 6.3 equiv.) was added to the reaction mixture and stirred for further 2 min, followed by the addition of propargyl bromide (0.53 mg, 4.47  $\mu$ mol, 3.4 equiv.). The reaction mixture was stirred at room temperature overnight. Upon completion of the product conversion, monitored *via* analytical RP-HPLC, the solvent was evaporated and the residue was dissolved in 10 mL H<sub>2</sub>O, filtered and purified two times *via* RP-HPLC. The triethylammonium salt of **8** which was further transferred to its sodium salt (general procedure 4) and isolated as a colourless solid after lyophilization (96%).

**<sup>1</sup>H NMR** (500 MHz, D<sub>2</sub>O)  $\delta$  8.50 (s, 1H, H2), 8.21 (s, 1H, H8), 6.11 (d,  $J$  = 6.0 Hz, 1H, H1'), 4.76 (t,  $J$  = 5.5 Hz, 1H, H2'), 4.60 – 4.50 (m, 1H, H3'), 4.40 – 4.34 (m, 1H, H4'), 4.32 – 4.14 (m, 2H, H5'), 3.51 (d,  $J$  = 12.4 Hz, 2H, CH<sub>2</sub>-S-P). **<sup>13</sup>C {<sup>1</sup>H} NMR** (201 MHz, MeOD)  $\delta$  154.26, 151.43, 147.74, 138.43, 117.28, 85.30, 82.55, 72.88, 68.98, 63.85, 21.82, 16.51, 16.49. **<sup>31</sup>P {<sup>1</sup>H} NMR** (202 MHz, D<sub>2</sub>O)  $\delta$  6.66 (d,  $J$  = 26.3 Hz, 1P,  $\gamma$ ), -11.56 (d,  $J$  = 19.5 Hz, 1P,  $\alpha$ ), -23.87 (dd,  $J$  = 26.4, 19.4 Hz, 1P,  $\alpha$ ).

**HRMS (ESI):** calculated for C<sub>13</sub>H<sub>15</sub>N<sub>5</sub>Na<sub>2</sub>O<sub>12</sub>P<sub>3</sub>S<sup>-</sup> [M-3H+2Na]<sup>-</sup>: 603.9452; found: 603.9453

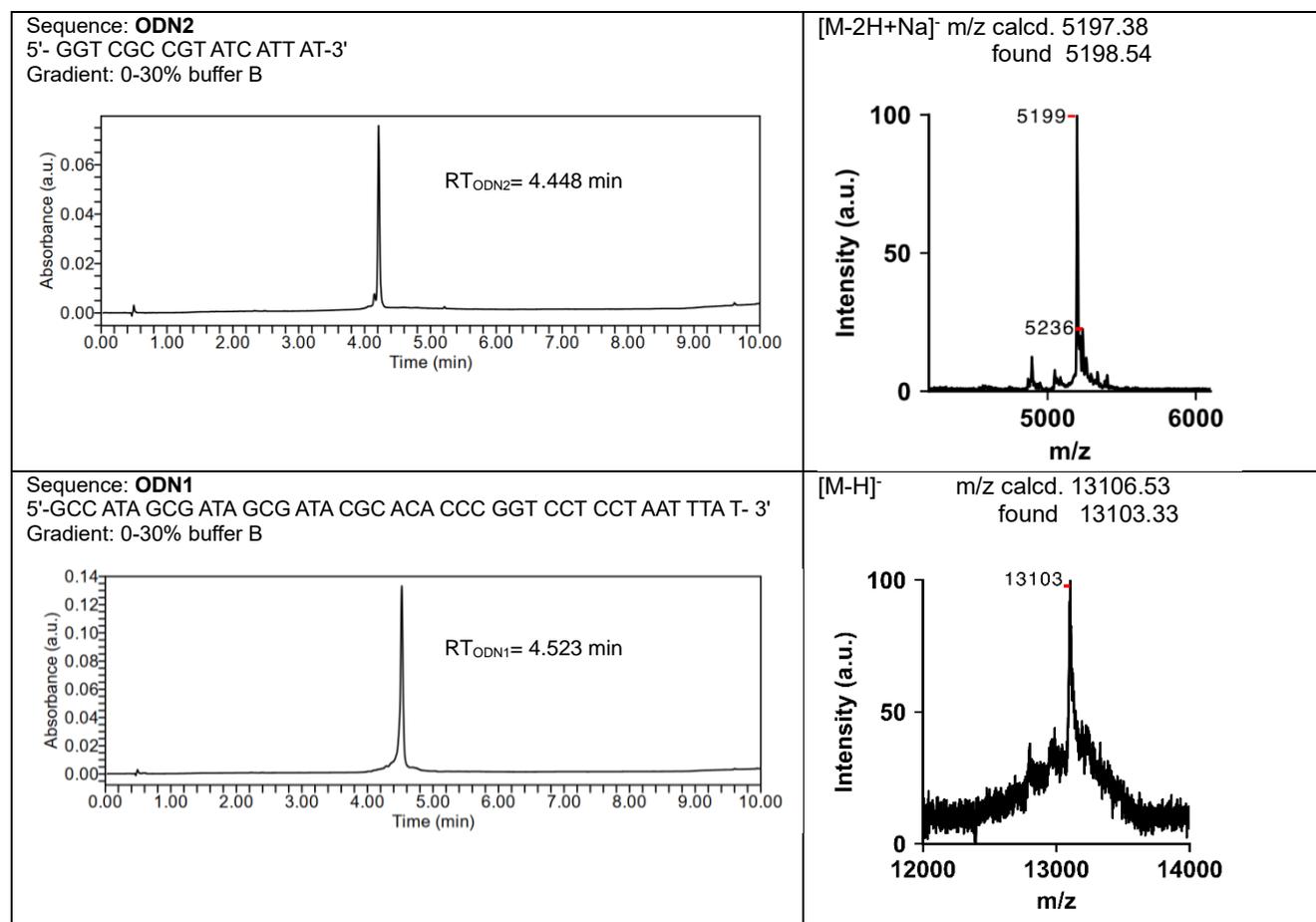
**Analytical RP-HPLC** (gradient: 0-10% Buffer B) R<sub>t</sub> = 4.821 min

## SUPPORTING INFORMATION

### 3.3 Synthesis and Characterization of unmodified DNA Oligonucleotides

Unmodified oligonucleotides were synthesised on 1  $\mu\text{mol}$  scale using DNA SynBase™ CPG 1000/110 (purchased from Link Technologies Ltd) as solid supports on an automated DNA synthesizer (K&A H-8 SE Oligo-Synthesizer) with standard phosphoramidite chemistry. Phosphoramidites (DMT-dA(bz), DMT-dC(ac), dmf-dG-CE, DMT-dT) were purchased from LinkTech and Sigma-Aldrich. The coupling efficiency was monitored with a trityl monitor. For the synthesis, DCA in  $\text{CH}_2\text{Cl}_2$  was employed as deblocking agent, Activator42® as an activator,  $\text{Ac}_2\text{O}$  in pyridine/THF as capping reagent and  $\text{I}_2$  as oxidizer in pyridine/ $\text{H}_2\text{O}$ . The cleavage and deprotection of the CPG bound oligonucleotides was carried out by treating the solid support with 1 mL of AMA solution (30%  $\text{NH}_4\text{OH}$ / 40% aqueous  $\text{MeNH}_2$ , 1:1 v/v). The suspension was heated on a thermoshaker for 20 min at  $65^\circ\text{C}$  and subsequently the supernatant was collected, and the beads were washed with water (3x0.5 mL). The combined aqueous solutions were concentrated under reduced pressure at  $30^\circ\text{C}$  using a SpeedVac concentrator (Savant™ SpeedVac™ SPD120 Vacuum Concentrator) and the remaining residue was lyophilized (CHRIST alpha 2-4 LD). Further, the oligonucleotides were purified by semi-preparative RP-HPLC as described in section 3.1, where a gradient of 0-50% of buffer B over 45 min was applied. The purified oligonucleotides were analysed by analytical RP-HPLC, as described in section 3.1, where a gradient of 0-30% buffer B from 0  $\rightarrow$  8 min, followed by 30-85% buffer B from 8  $\rightarrow$  10 min was applied. The absorbance of the oligonucleotides was measured at 260 nm using an Implen NanoPhotometer® P300. The extinction coefficients ( $\epsilon$ ) of the oligonucleotides were calculated using the OligoAnalyzer™ Tool from Integrated DNA Technologies. The concentrations of the oligonucleotides were calculated *via* Beer-Lambert law. The structural integrity of the synthesised oligonucleotides was determined by MALDI-TOF mass measurements as described in section 3.1.

**Table S7.** left) HPL-chromatograms. right) MALDI-TOF mass spectra (negative mode) of the purified oligonucleotides.



## SUPPORTING INFORMATION

### 3.4 T4 PNK catalysed 5'-end labelling reactions

5'-end labelling *via* T4 PNK without PEG: DNA or RNA oligonucleotides (5  $\mu$ L, 10  $\mu$ M, 50 pmol, 1 equiv.) in RNase free H<sub>2</sub>O, were heat denatured at 70 °C for 5 min and then immediately put on ice. After 5 min of cooling, T4 PNK buffer (5  $\mu$ L, 10X reaction buffer A, purchased from Thermo Scientific™),  $\gamma$ -modified ATP analogue (5  $\mu$ L, 10 mM in H<sub>2</sub>O, 50 nmol, 1000 equiv.) and T4 PNK (5  $\mu$ L, 10U/ $\mu$ L, purchased from Thermo Scientific™) were added. The reaction mixture (50  $\mu$ L total volume) was mixed by pipetting and incubated at 37°C for 3 h.

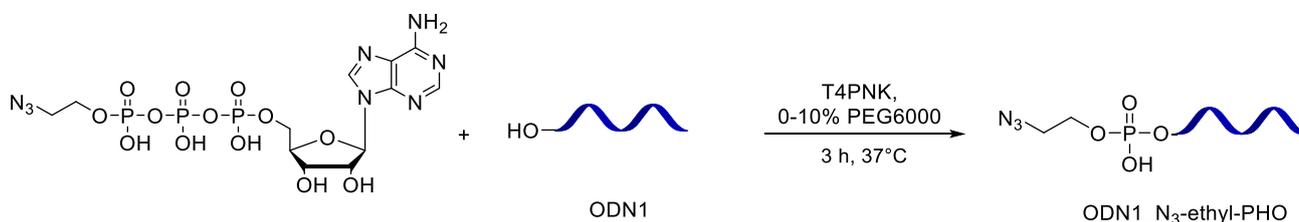
Purifications of T4 PNK reactions were performed using the Oligo Clean & Concentrator Kit from Zymo Research. The concentrations of the purified samples were determined on a Qubit 4 Fluorometer using the Qubit™ ssDNA Assay-Kit.

#### 3.4.1 Quantification of T4 PNK catalysed labelling reactions via analytical RP-HPLC

Analytical RP-HPLC was performed as described in section 3.1 and for consistency of analysis always one 50  $\mu$ L purified T4 PNK reaction (50 pmol of oligonucleotide) was injected. For quantification the peaks of unmodified and labelled oligonucleotide were integrated. The absorbance maximum for unmodified and 5'-end labelled oligonucleotide was assumed to be identical, for the purposes of this investigation. It was taken into account that the unmodified and 5'-labelled oligonucleotides have the same extinction coefficients ( $\epsilon$ ), since additional  $\gamma$ -modified phosphates (PHO) are not expected to substantially alter the absorption properties of 5'-end labelled oligonucleotides.

#### 3.4.2 Impact of PEG on substrate conversion

For analysis of the impact of PEG6000 as molecular crowding agent varying concentrations (0, 2.5, 5, 7.5, 10% in total) of PEG6000 were added as 24% (w/v) solution (purchased from Thermo Scientific™) to the 5'-end labelling reactions. Therefore, ODN1 (5  $\mu$ L, 10  $\mu$ M, 50 pmol) in RNase free H<sub>2</sub>O, was heat denatured at 70 °C for 5 min and then immediately put on ice. After 5 min of cooling, PEG6000 (0, 5.2, 10.4, 15.6 or 20.8  $\mu$ L of 24% solution), T4 PNK buffer (10X reaction buffer A), compound 9 (5  $\mu$ L, 10 mM in H<sub>2</sub>O, 50 nmol, 1000 equiv., purchased from Jena Bioscience) and T4 PNK (5  $\mu$ L, 10U/ $\mu$ L) were added. The reaction mixtures (50  $\mu$ L total volume) were mixed by pipetting and incubated at 37°C for 3 h. The purifications and subsequent analytical RP-HPLC analysis were performed as described in section 3.1. The results and relevant RP-HPL chromatograms can be taken from Table S8. The T4 PNK catalysed 5'-end labelling reaction is displayed in Scheme S1.



**Scheme S1.** T4 PNK catalysed 5'-azido labelling of an oligodeoxyribonucleotide (ODN1).

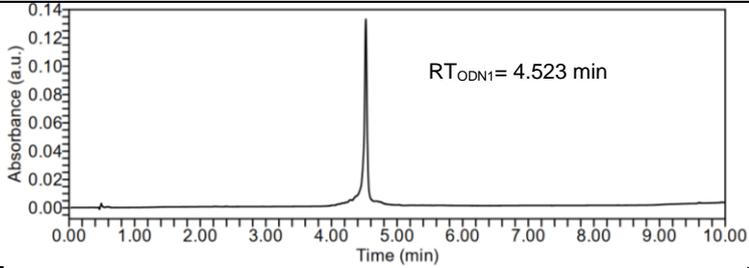
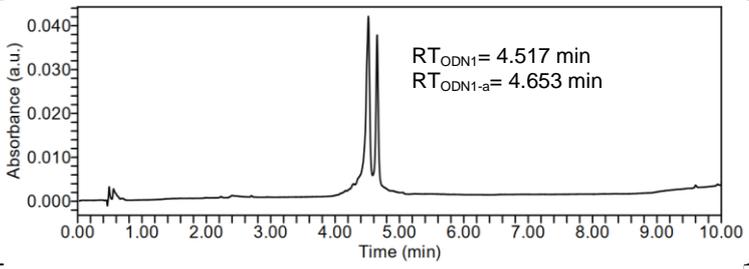
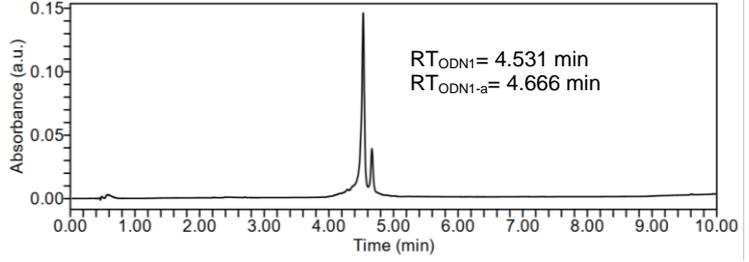
## SUPPORTING INFORMATION

**Table S8.** Impact of PEG6000 [%] on substrate conversion.  $RT_{ODN1\_N3\text{-ethyl-}PHO} = RT_{ODN1\text{-}a}$

Amount of PEG6000 in 50 $\mu$ L reaction [%]	Conversion yield of 5'-end labelling of ODN1 with compound 9 [%]	Analytical RP-HPL chromatogram (gradient: 0-30% buffer B)
0	16.64	<p> <math>RT_{ODN1} = 4.674 \text{ min}</math>  <math>RT_{ODN1\text{-}a} = 4.819 \text{ min}</math> </p>
2.5	30.90	<p> <math>RT_{ODN1} = 4.646 \text{ min}</math>  <math>RT_{ODN1\text{-}a} = 4.804 \text{ min}</math> </p>
5	36.82	<p> <math>RT_{ODN1} = 4.517 \text{ min}</math>  <math>RT_{ODN1\text{-}a} = 4.653 \text{ min}</math> </p>
7.5	33.39	<p> <math>RT_{ODN1} = 4.633 \text{ min}</math>  <math>RT_{ODN1\text{-}a} = 4.789 \text{ min}</math> </p>
10	35.55	<p> <math>RT_{ODN1} = 4.657 \text{ min}</math>  <math>RT_{ODN1\text{-}a} = 4.806 \text{ min}</math> </p>

## SUPPORTING INFORMATION

**Table S9.** As proof of concept for the peak assignment, 5  $\mu$ L of 10  $\mu$ M ODN1 were spiked to a the purified 50  $\mu$ L T4 PNK reaction, using compound 9 as substrate, with 5% PEG6000. Analytical RP-HPLC analysis as proof of concept for peak assignment.  $RT_{\text{ODN1-N3-ethyl-PhO=ODN1-a}}$

sample	Analytical RP-HPLC chromatogram (gradient: 0-30% buffer B)
unmodified ODN1	 <p><math>RT_{\text{ODN1}} = 4.523</math> min</p>
T4 PNK with compound 9 on ODN1 with 5% PEG (purified 50 $\mu$ L reaction)	 <p><math>RT_{\text{ODN1}} = 4.517</math> min <math>RT_{\text{ODN1-a}} = 4.653</math> min</p>
T4 PNK with compound 9 on ODN1 with 5% PEG (purified 50 $\mu$ L reaction) + 5 $\mu$ L of 10 mM ODN1 spiked	 <p><math>RT_{\text{ODN1}} = 4.531</math> min <math>RT_{\text{ODN1-a}} = 4.666</math> min</p>

### 3.4.3 Screening of $\gamma$ -phosphate modified ATP analogues

For the screening of  $\gamma$ -phosphate modified ATP analogues, the protocol for the T4 PNK catalysed 5'-end labelling reaction described in section 3.4.2 with 5% PEG6000 solution (10.2  $\mu$ L of 24% solution) and ODN1 was applied for all substrates (compounds 2-10). The purification and subsequent RP-HPLC analysis was performed as described in section 5 (Table S3). Compounds 3, 5, 6, 7 and 8 were synthesised as described in section 2, compounds 2 and 4 were either synthesised or purchased from Jena Bioscience and compounds 9 and 10 were purchased from Jena Bioscience.

## SUPPORTING INFORMATION

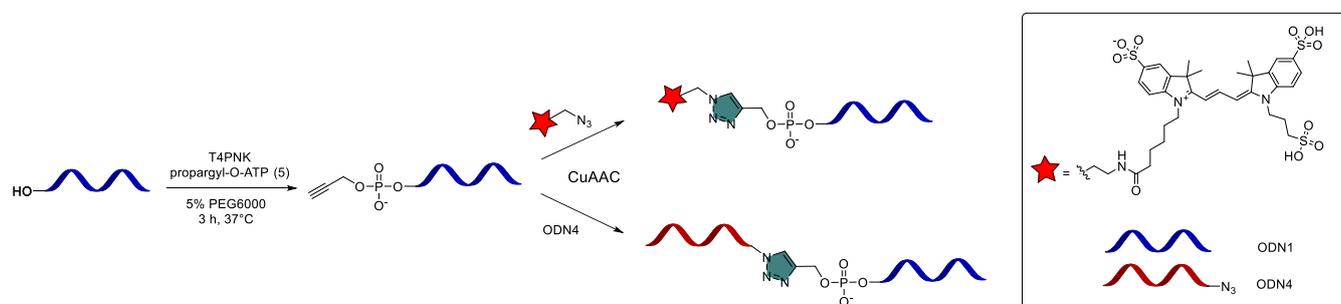
### 3.5 Click reactions with 5'-end labelled oligonucleotides

#### 3.5.1 Cu(I) catalysed azide-alkyne cycloaddition (CuAAC)

The T4 PNK catalysed 5'-end labelling reaction with propargyl-O-ATP (compound 4) as substrate was performed as described in section 3.4.2 with ODN1 (50 pmol). The purification and determination of concentration were performed as described in section 3.1. The Cu(I) source for the CuAAC was either a heterogeneous reactor catalyst system (baseclick GmbH) (**I**) or  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$  (**II**).

**Oligo-dye CuAAC: I**) CuAAC between 5'-end propargyl-PHO labelled ODN1 and a fluorescent dye (Scheme S2) was prepared by mixing the following components in a 1.5 mL tube: 5'-end propargyl-PHO labelled ODN1 (15  $\mu\text{L}$ , 37 pmol, 1 equiv.) was mixed with 3  $\mu\text{L}$  of activator solution (10X: 8 mM THPTA, 200 mM  $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$  in 50% DMSO/50%  $\text{H}_2\text{O}$  (v/v)) and 3.7  $\mu\text{L}$  of Sulfo-Cy3-Azide (1 mM in  $\text{H}_2\text{O}$ , 3.7 nmol, 100 equiv., purchased from Jena Bioscience). Afterwards, 2 pellets of the heterogeneous catalyst and 8.3  $\mu\text{L}$  nuclease-free  $\text{H}_2\text{O}$ , to reach a final volume of 30  $\mu\text{L}$ , were added, mixed and the click reaction was performed on a thermomixer at 45°C for 1.5 h at 600 rpm. Afterwards the supernatant was removed, the catalyst pellets were washed two times with 20  $\mu\text{L}$  nuclease-free  $\text{H}_2\text{O}$ , combined with the supernatant and purified as described in section 3.1. The purified product was analysed *via* dPAGE (section 1) and imaged before and after staining with GelRed® Nucleic Acid Stain (Figure S1A). **II**) 5'-end propargyl-PHO labelled ODN1 (7  $\mu\text{L}$ , 116.5 pmol, 1 equiv.) was added to 22.1  $\mu\text{L}$  degassed catalytic solution, containing THPTA (16.1  $\mu\text{L}$  of 10 mM in  $\text{H}_2\text{O}$ , 161 nmol, 1382 equiv.), sodium ascorbate (3  $\mu\text{L}$  of 100 mM in  $\text{H}_2\text{O}$ , 300 nmol, 2575 equiv.) and  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$  (3  $\mu\text{L}$  of 10 mM in  $\text{H}_2\text{O}$ , 30.0 nmol, 257.5 equiv.) and mixed well. Then, Sulfo-Cy3-Azide (11.65  $\mu\text{L}$  of 5 mM in  $\text{H}_2\text{O}$ , 58.3 nmol, 500 equiv.) and 4.25  $\mu\text{L}$  to reach a total volume of 45  $\mu\text{L}$  was added and mixed again. The reaction mixture was stored at r.t. over night and purified as described in section 3.1. The purified product was analysed *via* dPAGE (section 1) and imaged before and after staining with GelRed® Nucleic Acid Stain (Figure S1B)

**Oligo-oligo CuAAC: I**) CuAAC between 5'-end propargyl-PHO labelled ODN1 and ODN4 (Scheme S2) was performed by mixing the following components in a 1.5 mL tube: 5'-end propargyl-PHO labelled ODN1 (15  $\mu\text{L}$ , 31.6 pmol, 1 equiv.) was mixed with 3  $\mu\text{L}$  of activator solution (10X: 8 mM THPTA, 200 mM  $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$  in 50% DMSO/50%  $\text{H}_2\text{O}$  (v/v)) and ODN4 (1.5  $\mu\text{L}$ , 100  $\mu\text{M}$ , 150 pmol, 5 equiv.). Afterwards 2 pellets of heterogeneous catalyst and 10.5  $\mu\text{L}$  nuclease-free  $\text{H}_2\text{O}$ , to reach a final volume of 30  $\mu\text{L}$ , were added, mixed and the click reaction was performed on a thermomixer at 45°C for 1.5 h at 600 rpm. Afterwards the supernatant was removed, the catalyst pellets were washed two times with 20  $\mu\text{L}$  nuclease-free  $\text{H}_2\text{O}$ , combined with the supernatant and purified as described in section 3.1. The purified product was analysed *via* dPAGE (section 3.1) and imaged after staining with GelRed® Nucleic Acid Stain (Figure S1A).



**Scheme S2.** 5'-end propargyl labelling and subsequent CuAAC of a deoxoligonucleotide (ODN1).

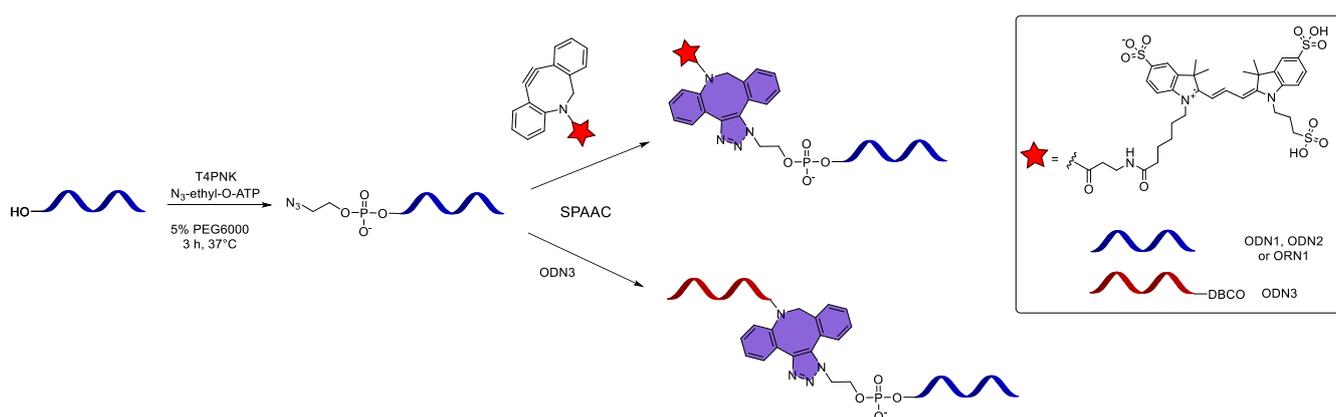
## SUPPORTING INFORMATION

### 3.5.2 Strain promoted azide-alkyne cycloaddition (SPAAC)

The T4 PNK catalysed 5'-end labelling reaction with N<sub>3</sub>-ethyl-O-ATP (compound 9) as substrate was performed as described in section 3.4.2 (5% PEG6000) with ODN1, ODN2 or ORN1 (50 pmol). The purification and determination of concentration were, unless otherwise specified, performed as described in section 3.1. For the Oligo-dye SPAAC the T4 PNK reaction of ODN1 and ODN2 was not purified (**C**), whereas for ORN1 the reaction mixture was purified as described above (**D**). The oligo-oligo SPAAC with ODN and ORNs, as well as oligo-dye SPAAC with ORN1 were performed following an adapted procedure based on the protocol from Taemaitree *et al.*<sup>[8]</sup> (Scheme S3).

**Oligo-dye SPAAC: C)** To 25  $\mu\text{L}$  of the crude T4 PNK reaction mixture containing ODN1 or ODN2 (20 pmol, 1 equiv.) and N<sub>3</sub>-ethyl-O-ATP (25 nmol, 1000 equiv.), 10  $\mu\text{L}$  of DBCO-Sulfo-Cy3 (5 mM in H<sub>2</sub>O, 50 nmol, 2000 equiv., purchased from Jena Bioscience) was added, mixed and stored at r.t. over night and purified as described in section 3.1. The purified product was analysed *via* dPAGE (section 1) and imaged before and after staining with GelRed® Nucleic Acid Stain (Figure S2A). **D)** 5'-end N<sub>3</sub>-ethyl-PHO labelled ORN1 (6.7  $\mu\text{L}$ , 23.5 pmol, 1 equiv.) was added to DBCO-Sulfo-Cy3 (4.7  $\mu\text{L}$ , 5 mM in H<sub>2</sub>O, 23.5 nmol, 1000 equiv.), HEPES solution (4.2  $\mu\text{L}$  of 0.1 M in H<sub>2</sub>O, pH 7-7.6), NaCl solution (15.8  $\mu\text{L}$  of 0.2 M in H<sub>2</sub>O), of EDTA solution (1.4  $\mu\text{L}$ , 0.1 M in H<sub>2</sub>O) and nuclease-free H<sub>2</sub>O (4.2  $\mu\text{L}$ ) to reach a total volume of 37  $\mu\text{L}$  and mixed well. The reaction mixture was stored at r.t. over night and purified as described in section 3.1. The purified product was analysed *via* dPAGE (section 3.1) and imaged before and after staining with GelRed® Nucleic Acid Stain (Figure S3).

**Oligo-oligo SPAAC:** 5'-end N<sub>3</sub>-ethyl-PHO labelled ODN1, ODN2 or ORN1 (23.5 pmol, 1 equiv.), was added to ODN3 (47 pmol, 2 equiv.), HEPES solution (4.2  $\mu\text{L}$  of 0.1 M in H<sub>2</sub>O, pH 7-7.6), NaCl solution (15.8  $\mu\text{L}$  of 0.2 M in H<sub>2</sub>O), EDTA solution (1.4  $\mu\text{L}$ , 0.1 M in H<sub>2</sub>O) and nuclease-free H<sub>2</sub>O to reach a total volume of 37  $\mu\text{L}$  and mixed well. The reaction mixture was stored at r.t. over night and purified as described in section 3.1. The purified product was analysed *via* dPAGE (section 3.1) and imaged after staining with GelRed® Nucleic Acid Stain (Figure S2B, S3).



**Scheme S3.** 5'-end azido labelling and subsequent SPAAC of an oligoribonucleotide (ORN1) or oligodeoxyribonucleotide (ODN1, ODN2).

## SUPPORTING INFORMATION

### 3.5.3 Dual labelling and subsequent click reactions

**(i) 3'- end labelling.** ODN2 (10  $\mu\text{L}$ , 10  $\mu\text{M}$ , 100 pmol, 1 equiv.) was added to 3'-azido-2',3'-ddGTP (2  $\mu\text{L}$ , 10 mM, 20 nmol, 200 equiv., provided by baseclick GmbH), TdT Reaction Buffer (10  $\mu\text{L}$ , 5X, purchased from Thermo Scientific™), TdT (4  $\mu\text{L}$ , 20 U/ $\mu\text{L}$ , Thermo Scientific™) and nuclease free  $\text{H}_2\text{O}$  (24  $\mu\text{L}$ ) to reach a total volume of 50  $\mu\text{L}$  and incubated at 37°C for 2 h. Then, the sample was purified, and the concentration was determined as described in section 3.1.

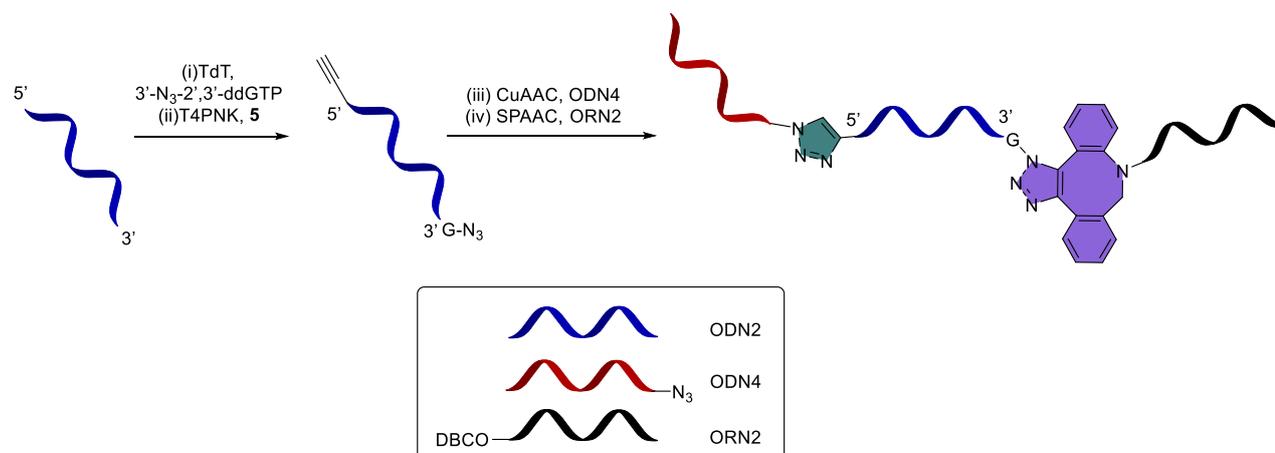
**(ii) 5'- end labelling.** ODN2-a (=3'- $\text{N}_3$ -dG-ODN2) (6.5  $\mu\text{L}$ , 50 pmol, 1 equiv.), from the previous 3'-end labelling **(i)**, was heat denatured at 70 °C for 5 min and then immediately put on ice. After 5 min of cooling, T4 PNK buffer (5  $\mu\text{L}$ , 10X reaction buffer A),  $\gamma$ -propargyl-O-ATP (5  $\mu\text{L}$ , 10 mM in  $\text{H}_2\text{O}$ , 50 nmol, 1000 equiv.), PEG6000 (10.4  $\mu\text{L}$ , 24% solution), T4 PNK (5  $\mu\text{L}$ , 10U/ $\mu\text{L}$ , purchased from Thermo Scientific™) and 18.1  $\mu\text{L}$  to reach a total volume of 50  $\mu\text{L}$ , were added. The reaction mixture was mixed by pipetting and incubated at 37°C for 3 h. The purification and determination of the concentration was performed as described in section 3.1.

**(iii) CuAAC on 5'- end.** Purified product from the previous step **(ii)** (18  $\mu\text{L}$ , 27.5 pmol, 1 equiv.) was mixed with 5  $\mu\text{L}$  of activator solution (10X: 8 mM THPTA, 200 mM  $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$  in 50% DMSO/50%  $\text{H}_2\text{O}$  (v/v)) and ODN4 (1.38  $\mu\text{L}$ , 200  $\mu\text{M}$ , 275 pmol, 10 equiv.). Afterwards 2 pellets of heterogeneous reactor catalyst system and nuclease-free  $\text{H}_2\text{O}$  (25.62  $\mu\text{L}$ ), to reach a final volume of 50  $\mu\text{L}$ , were added, mixed and the click reaction was performed on a thermomixer at 45°C for 1.5 h at 600 rpm. Afterwards the supernatant was removed, the catalyst pellets were washed two times with 20  $\mu\text{L}$  nuclease-free  $\text{H}_2\text{O}$ , combined with the supernatant and purified as described in section 3.1.

**(vi) SPAAC on 3'- end.** Purified product from the previous step **(iii)** (5  $\mu\text{L}$ , 80 pmol, 1 equiv.) was added to ORN2 (1.6  $\mu\text{L}$ , 100  $\mu\text{M}$ , 160 pmol, 2 equiv.), HEPES solution (4.2  $\mu\text{L}$  of 0.1 M in  $\text{H}_2\text{O}$ , pH 7-7.6), NaCl solution (15.8  $\mu\text{L}$  of 0.2 M in  $\text{H}_2\text{O}$ ), EDTA solution (1.4  $\mu\text{L}$ , 0.1 M in  $\text{H}_2\text{O}$ ) and nuclease-free  $\text{H}_2\text{O}$  (9  $\mu\text{L}$ ) to reach a total volume of 37  $\mu\text{L}$  and mixed well. The reaction mixture was stored at r.t. over night and purified as described in section 3.1.

The purified products of the individual steps was analysed *via* dPAGE (section 3.1) and imaged after staining with GelRed® Nucleic Acid Stain (Figure S4).

MALDI-TOF mass spectra of the individual labelling reactions, unmodified ODN2 and dually labelled ODN2 (ODN2-b) can be taken from Table S5.



**Scheme S4.** Dual labelling of ODN2 and subsequent click reactions.

## SUPPORTING INFORMATION

---

### 3.5.4 Circularization of ODN5 *via* SPAAC

**(I) 3'-end labelling** of ODN5 was performed as described in section 3.5.3 **(i)**,

**(II) 5'-end labelling.** 3'-N<sub>3</sub>-dG-ODN5 (7.75  $\mu$ L, 50 pmol, 1 equiv.), from the previous 3'-end labelling **(I)**, was heat denatured at 70 °C for 5 min and then immediately put on ice. After 5 min of cooling, T4 PNK buffer (5  $\mu$ L, 10X reaction buffer A),  $\gamma$ -N<sub>3</sub>-ethyl-O-ATP (5  $\mu$ L, 10 mM in H<sub>2</sub>O, 50 nmol, 1000 equiv.), PEG6000 (10.4  $\mu$ L, 24% solution), T4 PNK (5  $\mu$ L, 10U/ $\mu$ L, purchased from Thermo Scientific™) and 16.85  $\mu$ L to reach a total volume of 50  $\mu$ L, were added. The reaction mixture was mixed by pipetting and incubated at 37°C for 3 h. The purification and determination of the concentration was performed as described in section 3.1.

**(III) Circularization *via* SPAAC.** Purified ODN5 from the previous step **(II)** (11.54  $\mu$ L, 40 pmol, 1 equiv.) was heated to 70°C for 7 min and then slowly cooled to room temperature. Then, DBCO-PEG4-DBCO **(12)** (0.6  $\mu$ L, 10 mM in in 50% DMSO/50% H<sub>2</sub>O (v/v), 6 nmol, 150 equiv., purchased from BroadPharm), HEPES solution (4.2  $\mu$ L of 0.1 M in H<sub>2</sub>O, pH 7-7.6), NaCl solution (15.8  $\mu$ L of 0.2 M in H<sub>2</sub>O), EDTA solution (1.4  $\mu$ L, 0.1 M in H<sub>2</sub>O) and nuclease-free H<sub>2</sub>O (3.46  $\mu$ L) to reach a total volume of 37  $\mu$ L and mixed well. The reaction mixture was stored at r.t. over night. The purification and determination of the concentration was performed as described in section 3.1.

**(IV) Digestion of linear and circular ODN5 with MspI restriction enzyme.** Unmodified ODN5 or purified, labelled ODN5 from the previous step **(III)** (1-6  $\mu$ L, 9.5 pmol, 1 equiv.) was added to ODN6 (9.5  $\mu$ L, 10  $\mu$ M, 95 pmol, 10 equiv.), Tango Buffer (2  $\mu$ L, 10X, purchased from Thermo Scientific™) and nuclease free H<sub>2</sub>O (to reach a volume of 18  $\mu$ L), mixed, heated to 85°C for 7 min and then cooled to 24°C with a ramp of 0.1°C/s. Then, MspI (2  $\mu$ L, 10U/ $\mu$ L, purchased from Thermo Scientific™) was added, mixed by pipetting and incubated at 37°C for 16 h. The purification and determination of the concentration was performed as described in section 5. The purified product was analysed *via* dPAGE (section 3.1) and imaged after staining with GelRed® Nucleic Acid Stain (Figure S5).

ODN 5 unmodified displays small impurities of lower molecular weight, from SPS, which were also confirmed *via* analytical HPLC. Thus, this fragments with a smaller molecular weight are present the lane of the subsequent reaction steps.

## SUPPORTING INFORMATION

### 4. T4 PNK-mediated analogue hydrolysis assay (ADP-Glo Assay)

EC<sub>50</sub> values for T4 PNK-mediated hydrolysis of ATP analogues (compound 2, 4, 8, 9) to ADP were measured as proxies for the affinity of T4 PNK to the ATP analogues. Relative ADP concentrations were determined with the ADP-Glo™ Kinase Assay kit (purchased from Promega catalogue no. V6930) following the optimised forward kinase reaction (section 5.3 with 5% PEG6000). Per test compound (ATP analogue) and twice for ATP, wells 2 – 12 of one respective row of a 96-well plate (white, conical bottom, Sarstedt, catalogue no. 72.1982.202) were filled with 15 µL each of the kinase reaction mix without T4 PNK (1 mM ATP analogue, 1 µM ODN1, 5% PEG 6000, 1X T4 PNK buffer A), keeping the plate on ice. Wells one of each row were then filled with 30 µL each of the respective forward kinase reaction mixes also containing 1 U/µL T4 PNK. 15 µL of each first well were then transferred to each second well and thoroughly mixed by pipetting with a multi-channel pipette (Eppendorf), going through the wells for 1:2 dilutions until wells 11 were reached. The remaining 15 µL were discarded, and wells 12 were left as no-PNK controls. Two times 7 µL of each well were transferred to adjacent rows of new plates, producing technical duplicates. ATP-kinase reactions were present on all plates as a positive control. The plates were then sealed with adhesive film (GeneON catalogue no. A26979) and incubated at 37°C for 3 h in a thermal cycler with a heated lid (Analytik Jena). The ADP-Glo assay was then performed as per the manufacturer's instructions. Briefly, 7 µL of ADP-Glo Reagent was added to each well and incubated at room temperature for 40 min. 14 µL of Kinase Detection Reagent was added to each well and incubated for 30 min at room temperature. The luminescence per well was finally determined using a Tecan Infinite M Plex plate reader with 250 ms integration time.

Luminescence values for the no-PNK controls were subtracted from all measurements of the same compound in preparation for curve fitting. Data were fitted to a simple 4-parameter logistic model (Origin 2022 software) according to the formula below. The data was lastly scaled to the upper and lower asymptote of each fit for ease of visualisation, where applicable.

The ATP analogues were present in a 1000-fold molecular excess to the ODN1 in these tests, and thus a relatively weak signal is expected for ADP stemming only from the labelling reaction, compared to a signal obtained due to unspecific compound hydrolysis to form ADP.

Given that we previously observed instability of the N-linked alkyne 2, we speculate that the behaviour of this molecule could mirror that of ATP in the labelling reaction following autohydrolysis. (Scheme S5). On the other hand, ODN-independent ATP analogue hydrolysis by T4 PNK appears to be blocked by stable conjugation of the γ-phosphate, which is, however, not indicative of successful labelling in the case of the S-linked alkyne 8. Adenosine 5'-(γ-thio)triphosphate (ATP-γS) is a commonly used kinase substrate for enzymatic protein labelling. It exhibits a greatly reduced specific constant with protein kinase A (PKA) compared to ATP.<sup>[9]</sup> The PKA mechanism of phosphorylation is also thought to involve direct nucleophilic attack of the target peptide serine –OH and in-line transfer of the phosphate<sup>[10]</sup>, not unlike T4 PNK-mediated phosphorylation. We, therefore, deduce that the sulphur atom itself interferes with the core mechanism of phosphotransfer. The substrate specificity of the T4 PNK for the differing ATP analogues is not immediately obvious from the detected labelling rates alone, given that analogues with mostly short linkers lead to successful labelling but long linkers and bulky modifications also worked in the case of azide **10** and ATP-biotin<sup>[11]</sup>, albeit to a lesser extent.

$$f(x) = A + \frac{B - A}{1 + \left(\frac{x}{C}\right)^D}$$

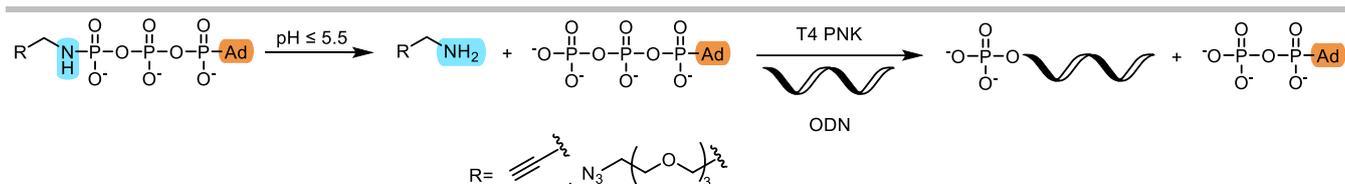
*A: Upper asymptote*

*B: Lower asymptote*

*C: Inflection point or EC<sub>50</sub>*

*D: Power or slope at inflection point*

## SUPPORTING INFORMATION



**Scheme S5.** Hypothesized mechanism for the phosphorylation of ODN with  $\gamma$ -N-linked ATP analogues.

## 5. Molecular docking analysis

Molecular docking analysis was performed based on the T4 PNK crystal structure available from PDB entry 1RRC.<sup>[12]</sup> The structure was prepared for docking by removing unbound atoms and ADP. Next, a magnesium ion was added to the ATP binding pit, with its position inferred from the bacterial PNK structure available from PDB entry 4MDE.<sup>[13]</sup> Polar hydrogens were added using AutoDock Tools (v1.5.7p1), and the structure was exported as a .pdbqt file. The file was then edited to assign a charge of +2 to the magnesium. Ligands were similarly prepared for docking using AutoDock Tools, adding polar hydrogens and auto-selecting rotatable bonds. AutoDock Vina<sup>[14],[15]</sup>, with the vinao scoring function<sup>[16]</sup>, was used for docking (grid centre: X 7.384 Y 85.473 Z 111.532, grid size: X 16 Y 22 Z 14, exhaustiveness: 100, random seed: 42). Docked poses in .pdbqt file format were converted to .sdf format using OpenBabel (v2.3.1, Chris Morley, openbabel.org). Poses were then visualised using py3Dmol (pypi.org/project/py3Dmol/) and RDKit (v2023.3.2, rdkit.org) for parsing.

The RMSD of the docked compounds  $\alpha$ - and  $\beta$ -phosphate C-O-P-O-P backbone to those of the ADP crystal structure pose was calculated using a custom Python script based on a solution described by Angle Ruiz-Moreno (<https://chem-workflows.com/articles/2019/06/24/pose-clustering-of-docking-results/>), except that the RMSD calculation was not limited to the maximum common substructure, but to the common C-O-P-O-P phosphate backbone substructure, anchoring the substructure search to the phosphate ester. The key change to the script is provided:

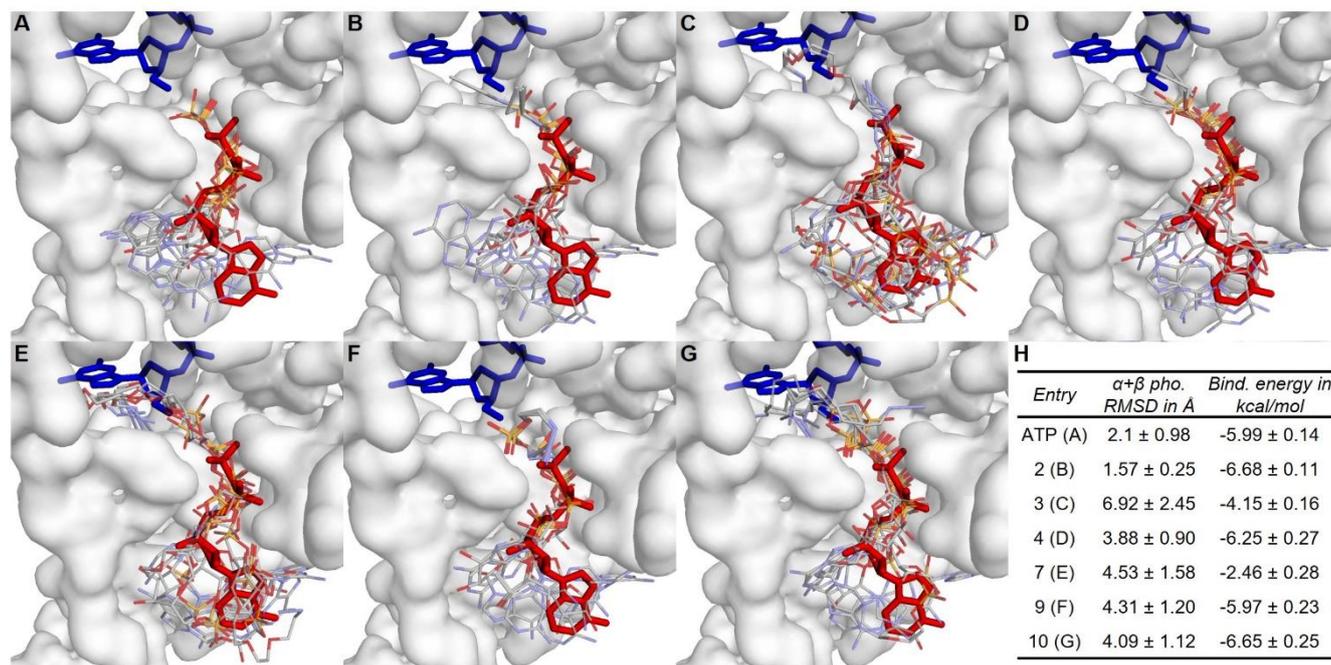
```
r=rdFMCS.FindMCS([mol,Chem.MolFromSmiles('POPOC')])
```



The docking results revealed a high positional variability for the adenosine moieties of all compounds, likely reflecting the known acceptance of T4 PNK towards different NTP substrates.<sup>[17]</sup> The phosphates of nearly all analogues and ATP itself as a control (Figure S6 A), on the other hand, were orderly positioned within the NTP binding pocket, and the  $\gamma$ -phosphates mostly extended beyond the  $\beta$ -phosphate of the crystallographically recorded ADP ligand by one unit length and were well positioned towards the ODN 5'-OH for nucleophilic attack and  $\gamma$ -phosphotransfer. In these cases,  $\gamma$ -modifications involving relatively hydrophobic alkyl linkers folded over and away from the enzyme active site (e.g., Figure S6 B – E), possibly facilitating the  $\gamma$ -phosphotransfer. Labelling was successfully detected in all these cases, with the exception of the N-linked alkyne 2, which was presumed to be unstable under test conditions. Labelling for long-linker azides 3 and 7, however, could not be detected. These compounds feature hydrophilic PEG linkers, and the docking analysis showed that these linkers could become embedded within the binding pit to compete with the polyphosphates, potentially disallowing the labelling reaction to occur (Figure S6 F, G) and offering a plausible explanation for the experimental data.

We, furthermore, aimed to extract predictive value from the docking simulation by calculating the difference in position of each compound's  $\alpha$ - and  $\beta$ -phosphates compared to those of the crystal structure ADP pose (Figure S6 H). Compared to ATP,  $\gamma$ -modifications seemed to influence the correct positioning of the phosphates, possibly explaining the reduced labelling efficiency. Indeed, labelling was the most efficient for alkyne 4, which also had the smallest difference in pose compared to ADP from the crystal structure of all analogues tested (except for allegedly unstable alkyne 2). Complementarily, binding energies directly provided by the docking software also seemed to be good predictors of the general possibility of labelling (Figure S6 H), although differences in pose seem to be more predictive of labelling efficiency overall. Together, these results may allow in silico prediction of labelling from novel compounds before the effort for their synthesis is made.

## SUPPORTING INFORMATION

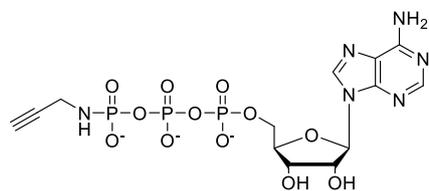


**Figure S6.** Rigid receptor docking of ATP analogues to the T4 PNK kinase active site based on PDB 1RRC. A-G: ATP binding pit of T4 PNK kinase active site (grey surface representation), bound ADP (red stick model) and bound dinucleotide (blue stick model) in their respective crystallographic poses, as well as nine docked poses per compound (thinner opaque stick models). H: Table of compounds per panel as well as RMSD in ångström of each compound's  $\alpha$  and  $\beta$  phosphates to the  $\alpha$  and  $\beta$  phosphates of the crystallographic ADP pose, as well as binding energies, including standard deviations. RMSD: Root Mean Squared Difference.

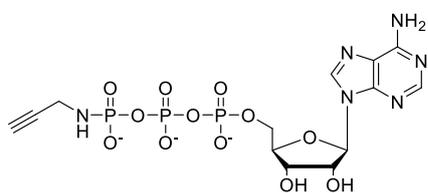
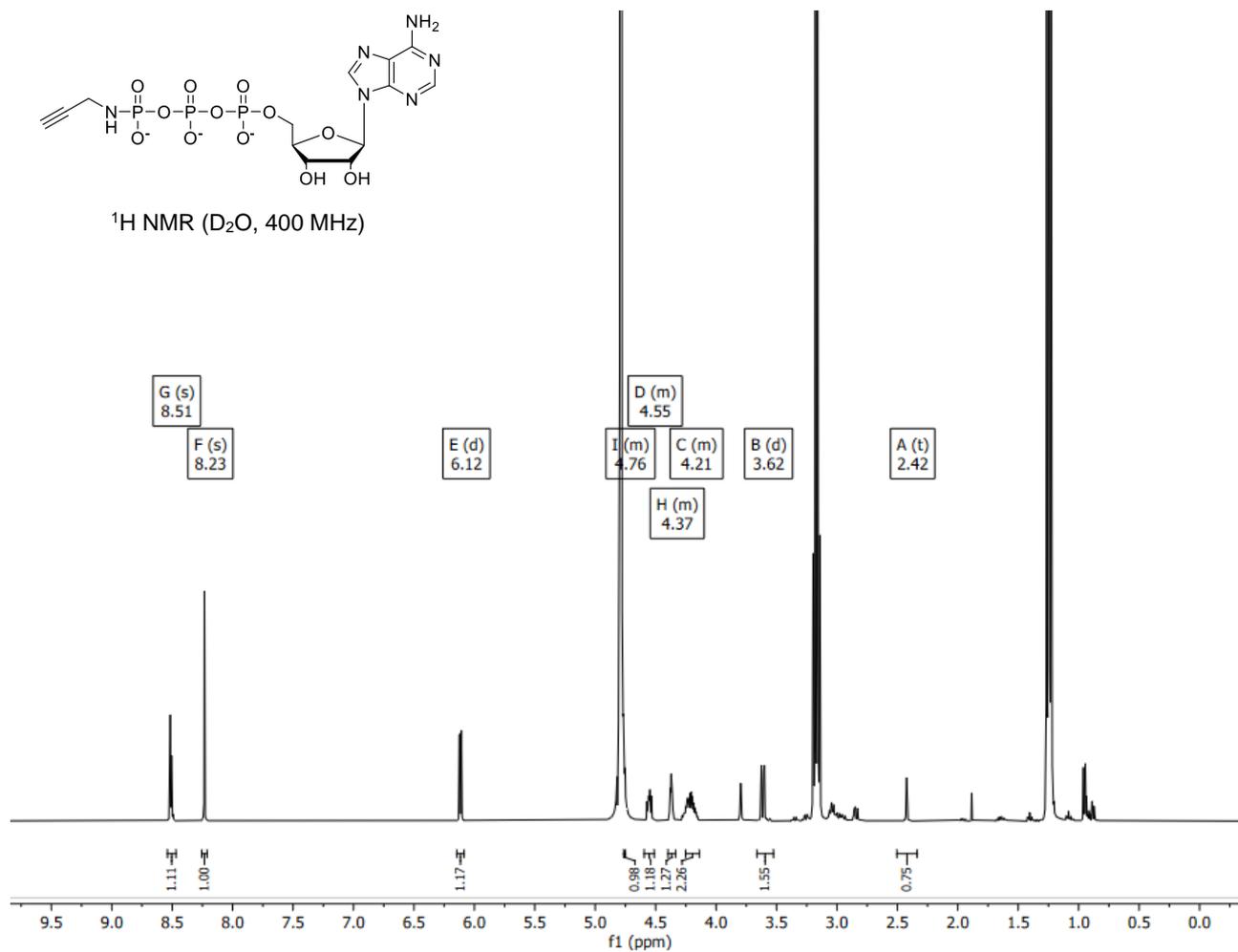
## SUPPORTING INFORMATION

### 6. NMR spectra of synthesised compounds

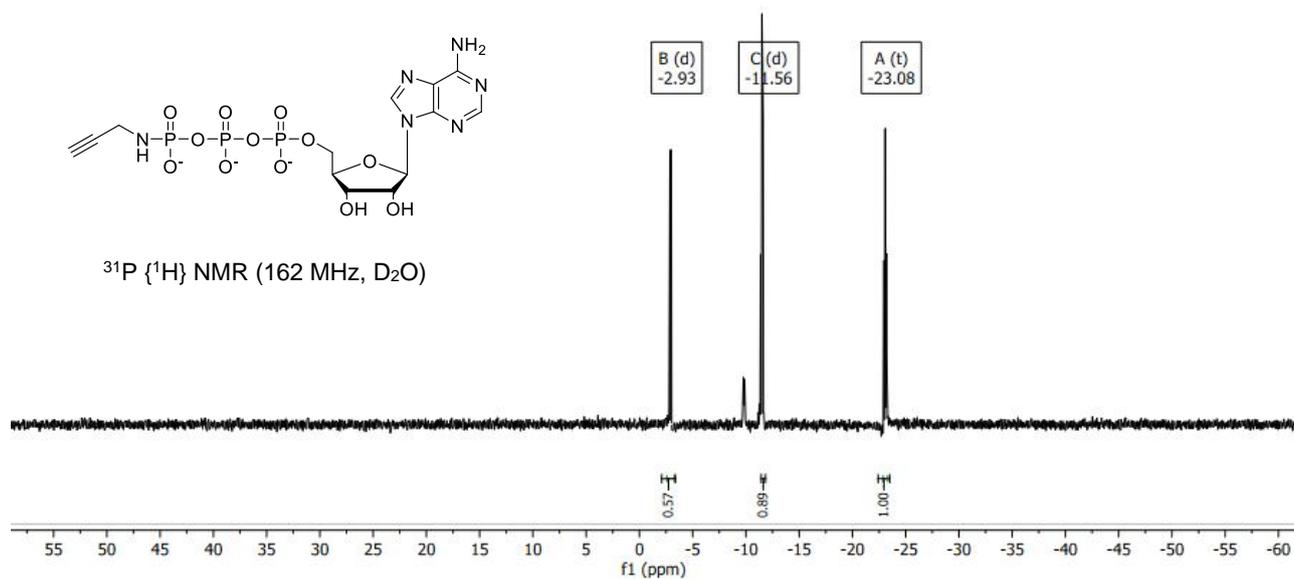
$^1\text{H}$ ,  $^{13}\text{C}$   $\{^1\text{H}\}$  and  $^{31}\text{P}$   $\{^1\text{H}\}$  spectra of compound 2



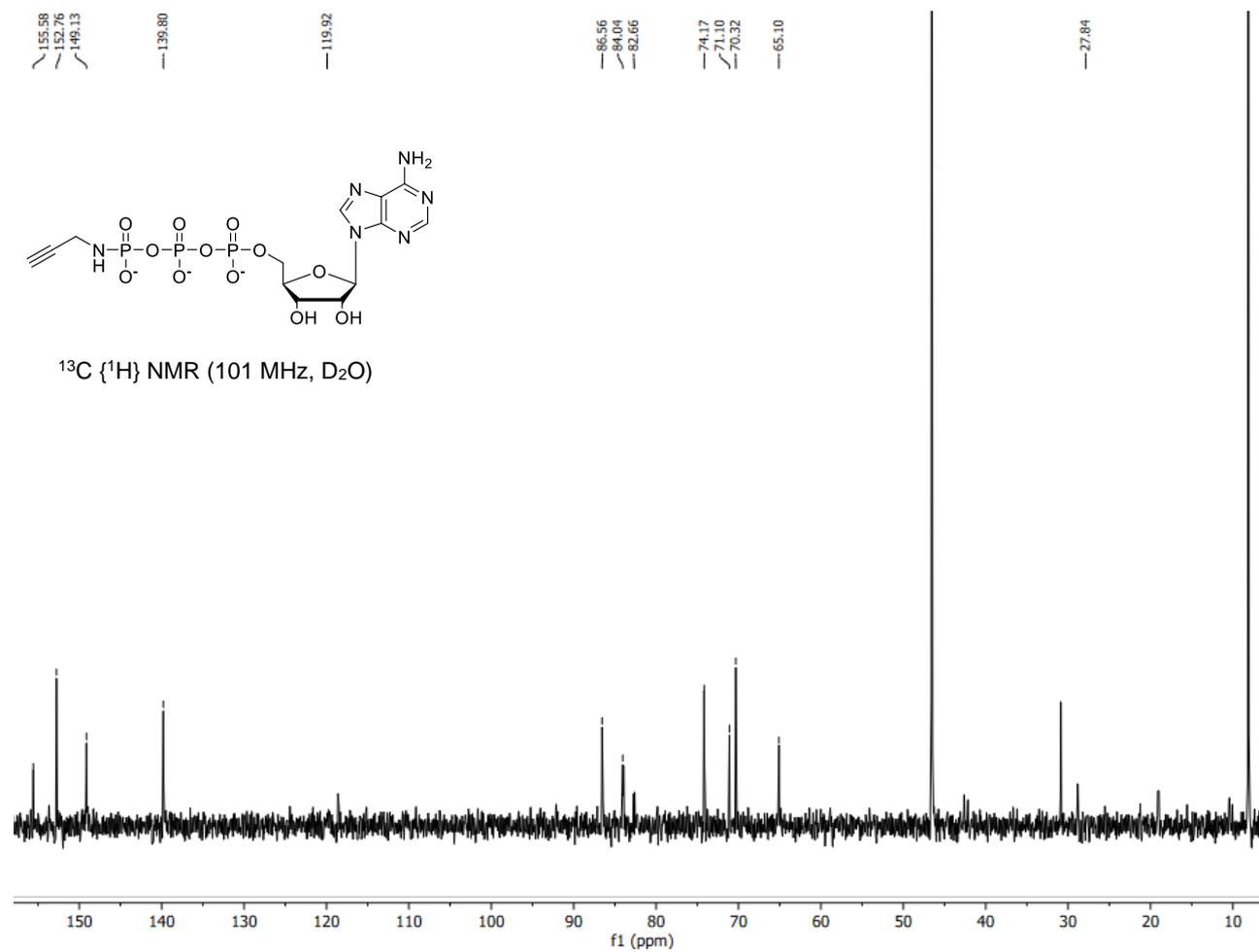
$^1\text{H}$  NMR ( $\text{D}_2\text{O}$ , 400 MHz)



$^{31}\text{P}$   $\{^1\text{H}\}$  NMR (162 MHz,  $\text{D}_2\text{O}$ )

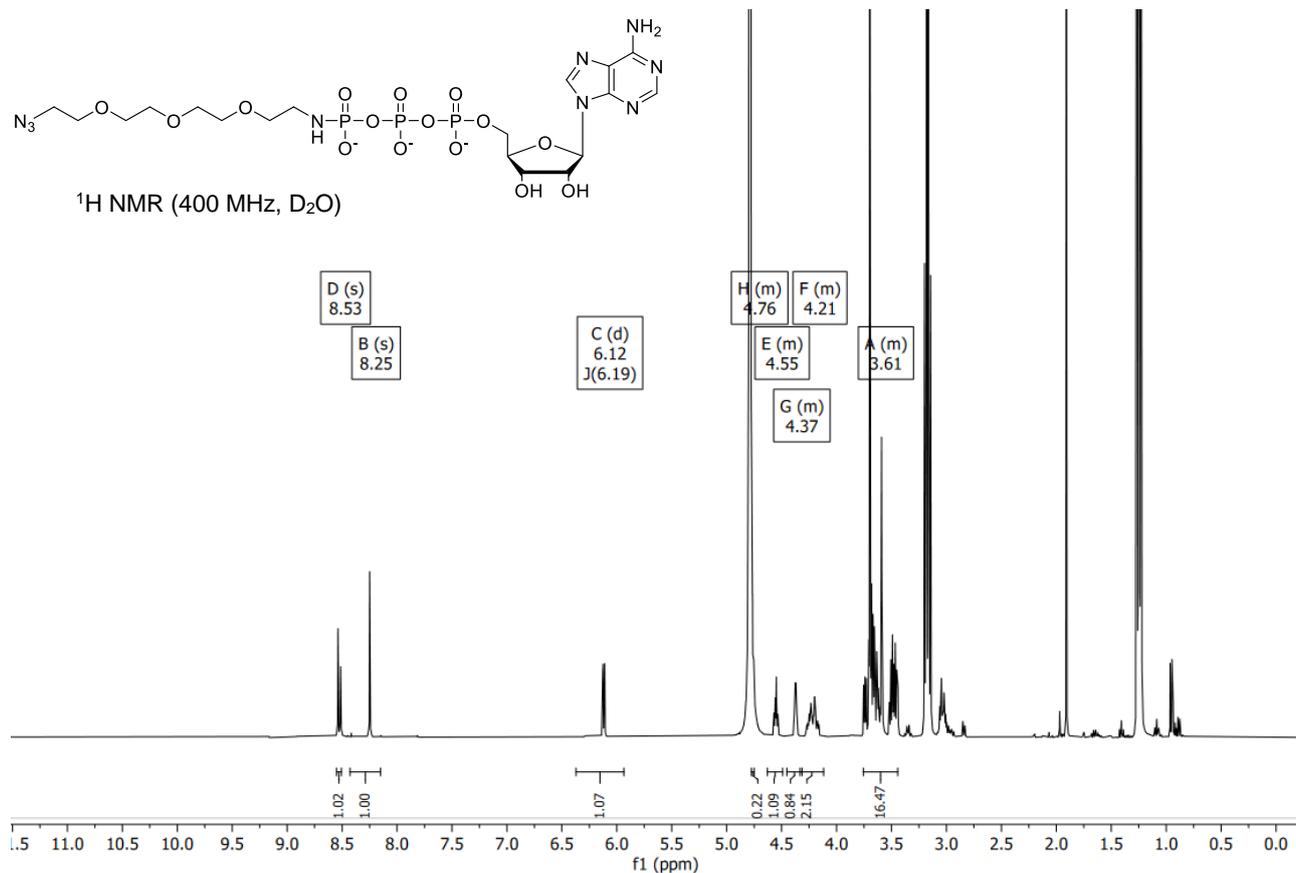


# SUPPORTING INFORMATION

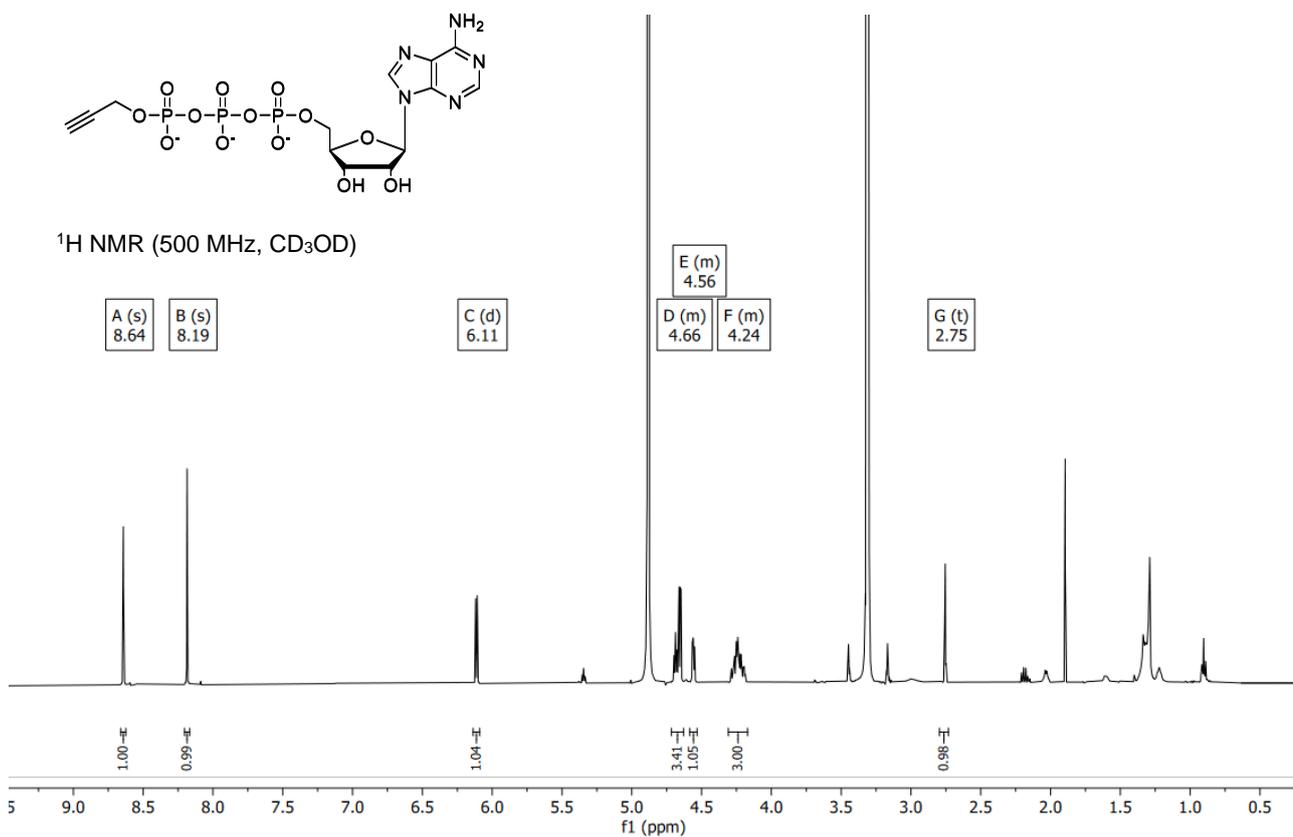


## SUPPORTING INFORMATION

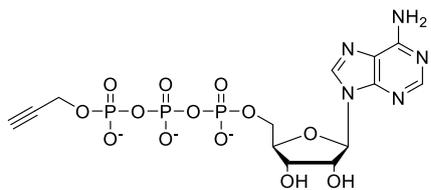
### $^1\text{H}$ spectrum of compound 3



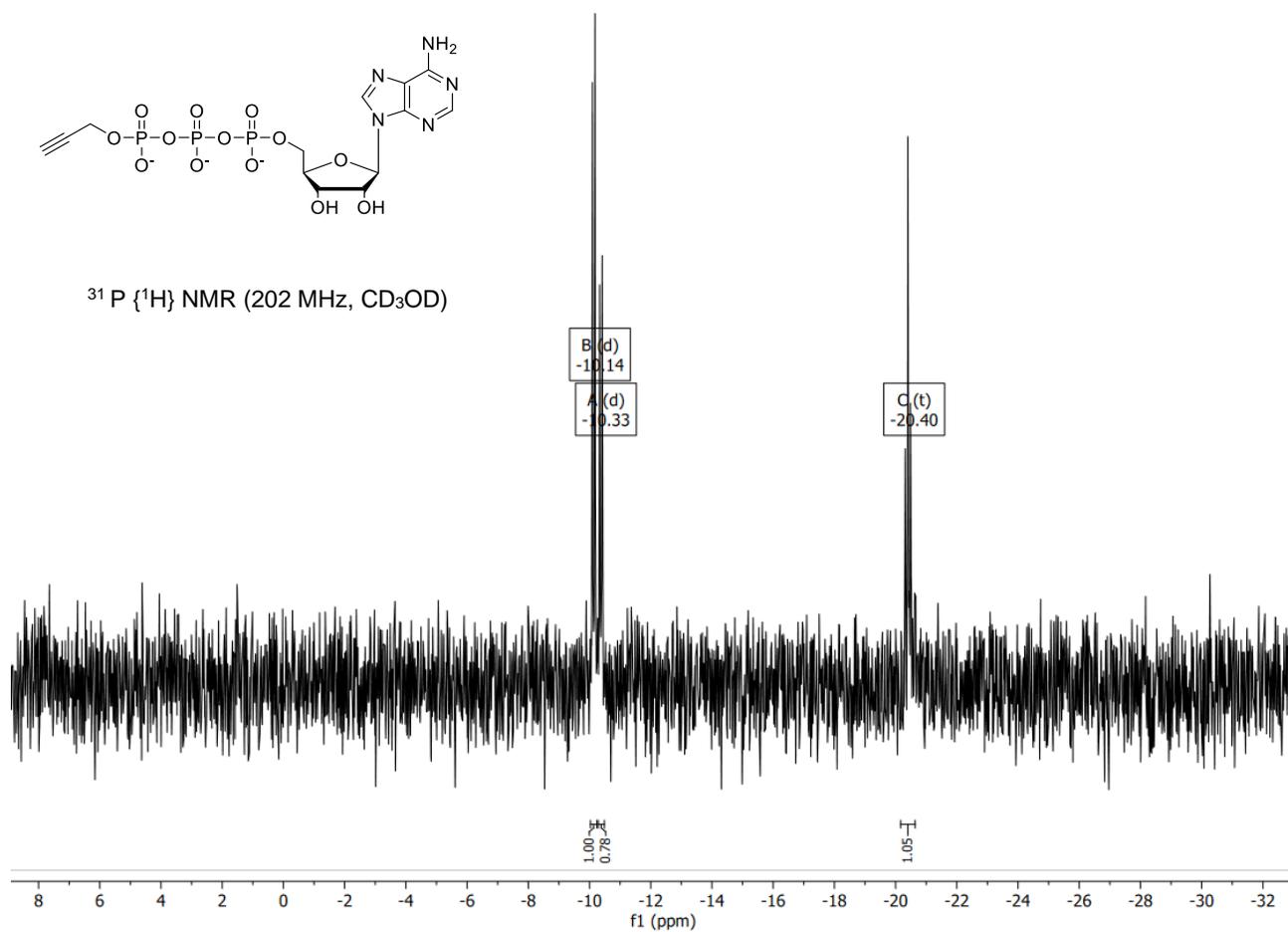
### $^1\text{H}$ , $^{31}\text{P}$ $\{^1\text{H}\}$ and HSQC spectra of compound 4



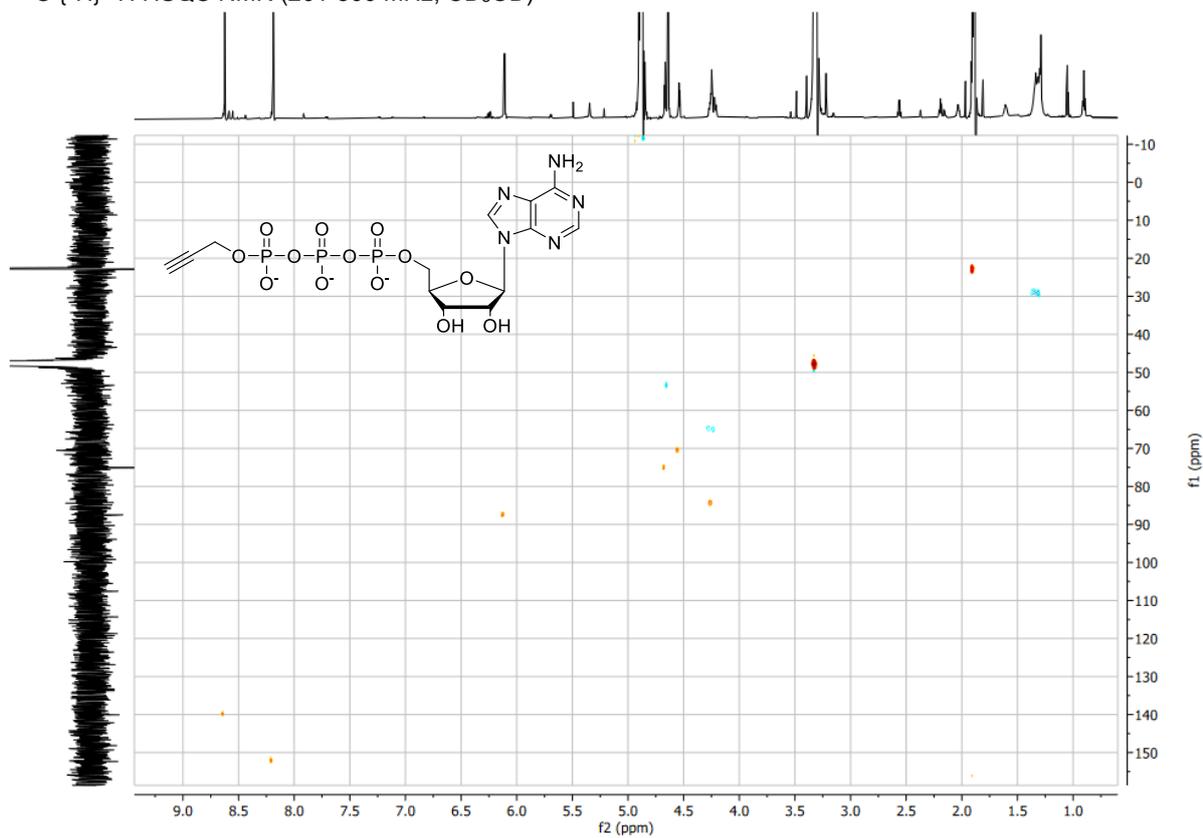
# SUPPORTING INFORMATION



$^{31}\text{P}$   $\{^1\text{H}\}$  NMR (202 MHz,  $\text{CD}_3\text{OD}$ )

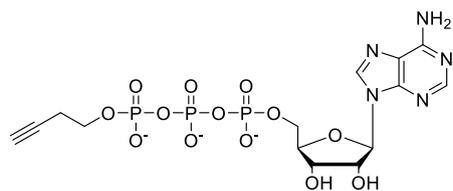


$^{13}\text{C}$   $\{^1\text{H}\}$ - $^1\text{H}$  HSQC NMR (201-500 MHz,  $\text{CD}_3\text{OD}$ )

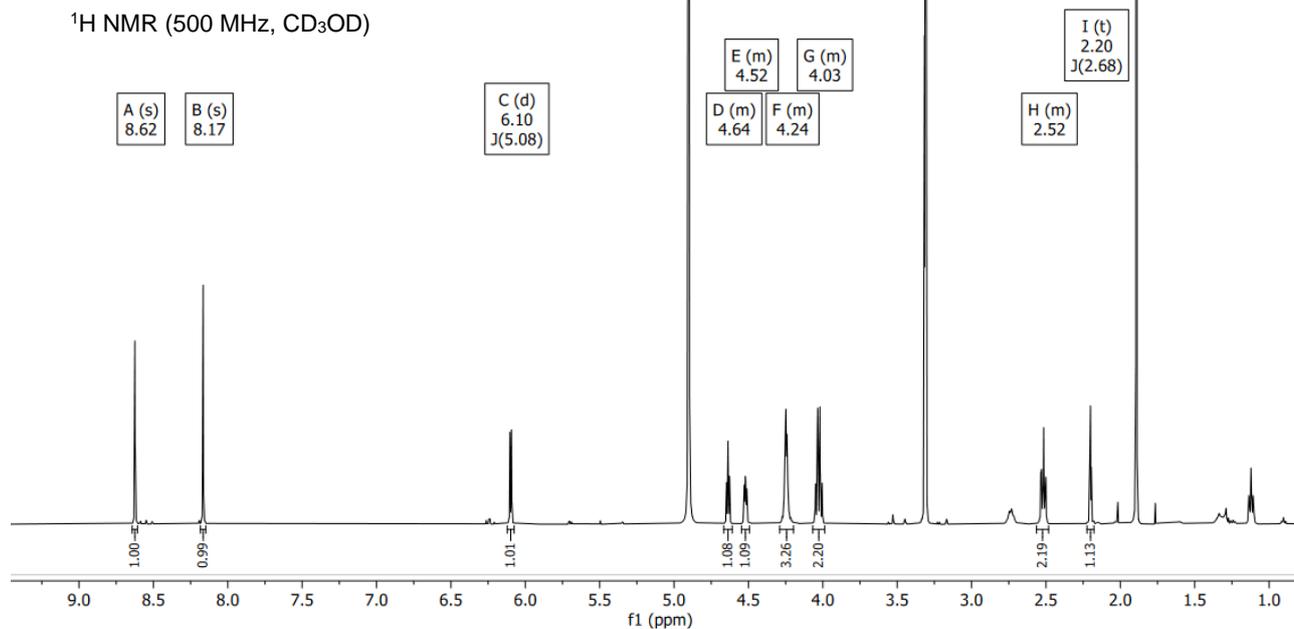


# SUPPORTING INFORMATION

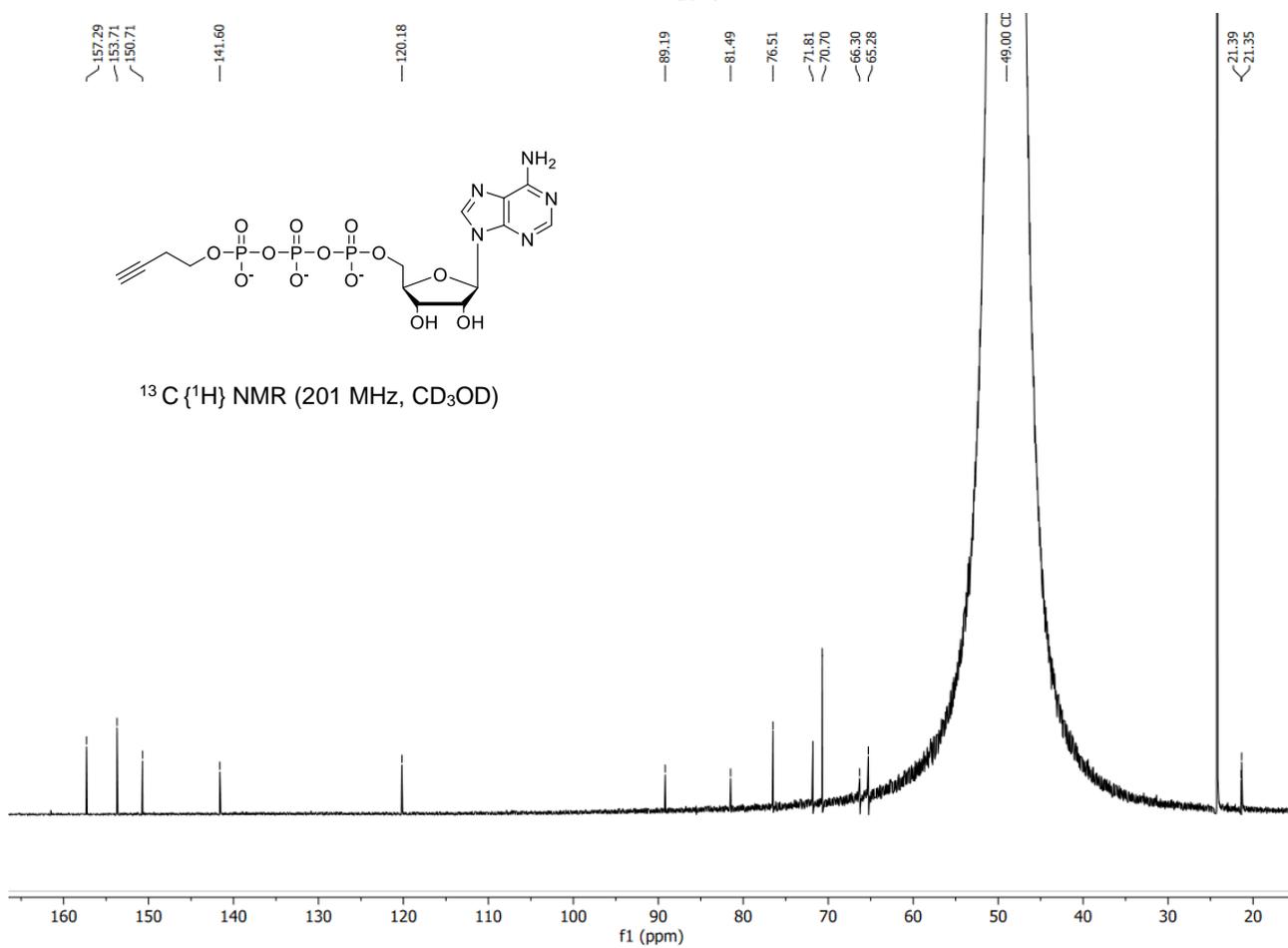
## $^1\text{H}$ , $^{13}\text{C}$ $\{^1\text{H}\}$ and $^{31}\text{P}$ $\{^1\text{H}\}$ spectra of compound 5



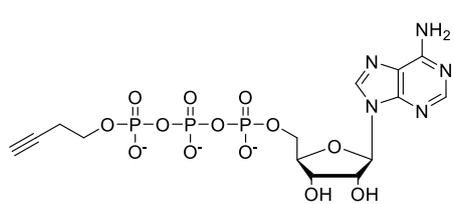
$^1\text{H}$  NMR (500 MHz,  $\text{CD}_3\text{OD}$ )



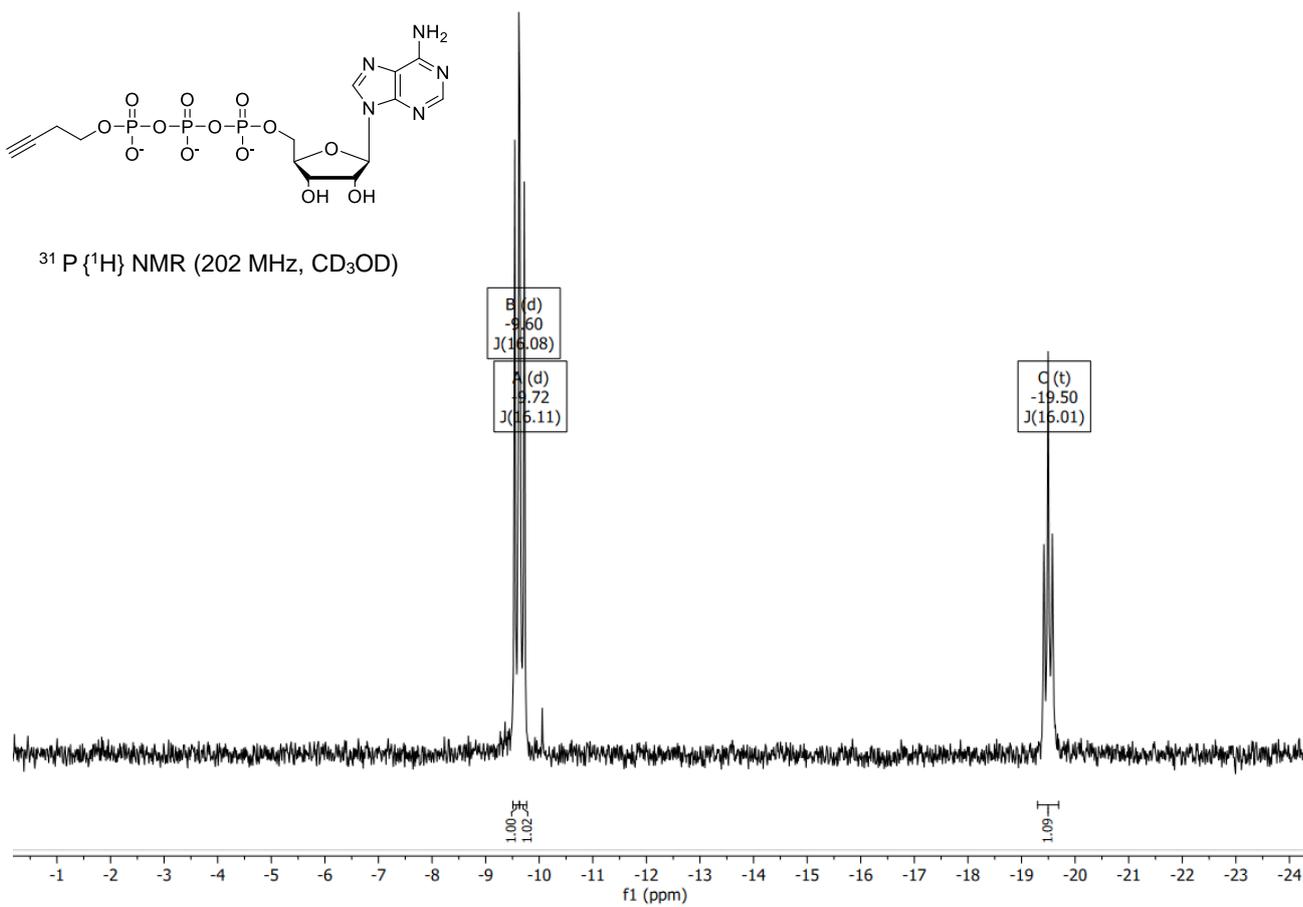
$^{13}\text{C}$   $\{^1\text{H}\}$  NMR (201 MHz,  $\text{CD}_3\text{OD}$ )



# SUPPORTING INFORMATION

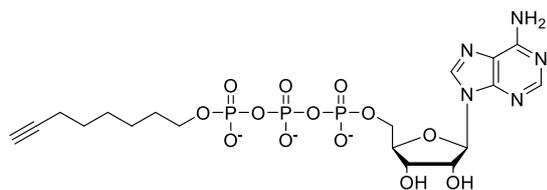


$^{31}\text{P}$  { $^1\text{H}$ } NMR (202 MHz,  $\text{CD}_3\text{OD}$ )

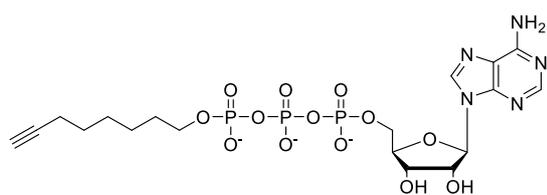
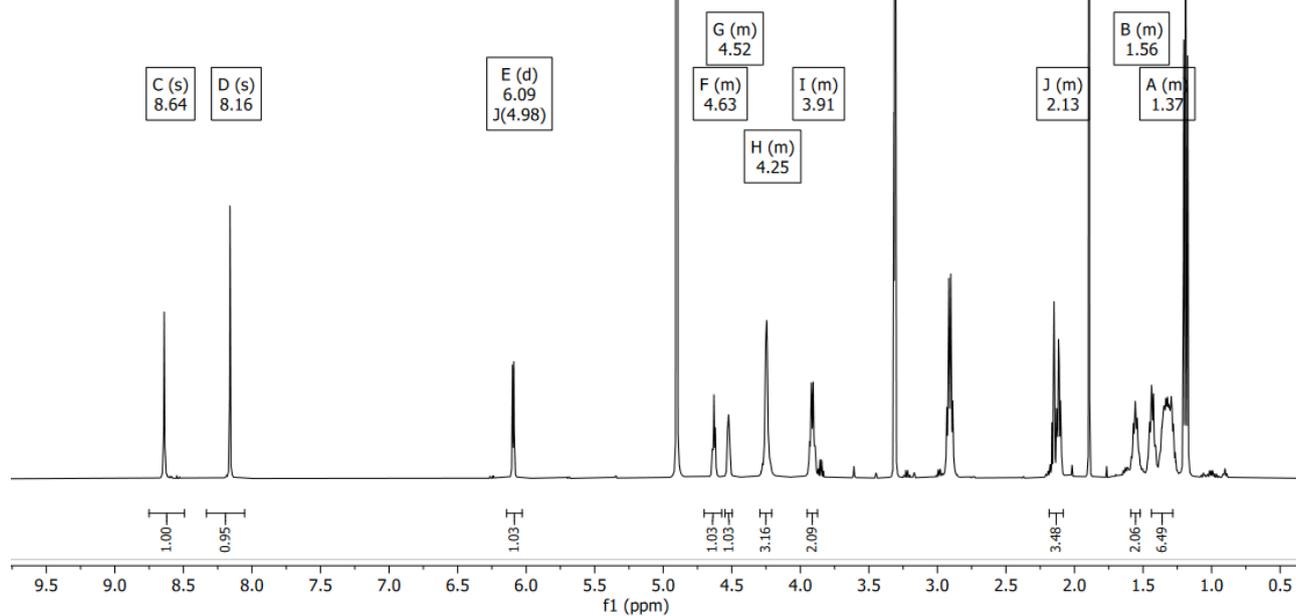


# SUPPORTING INFORMATION

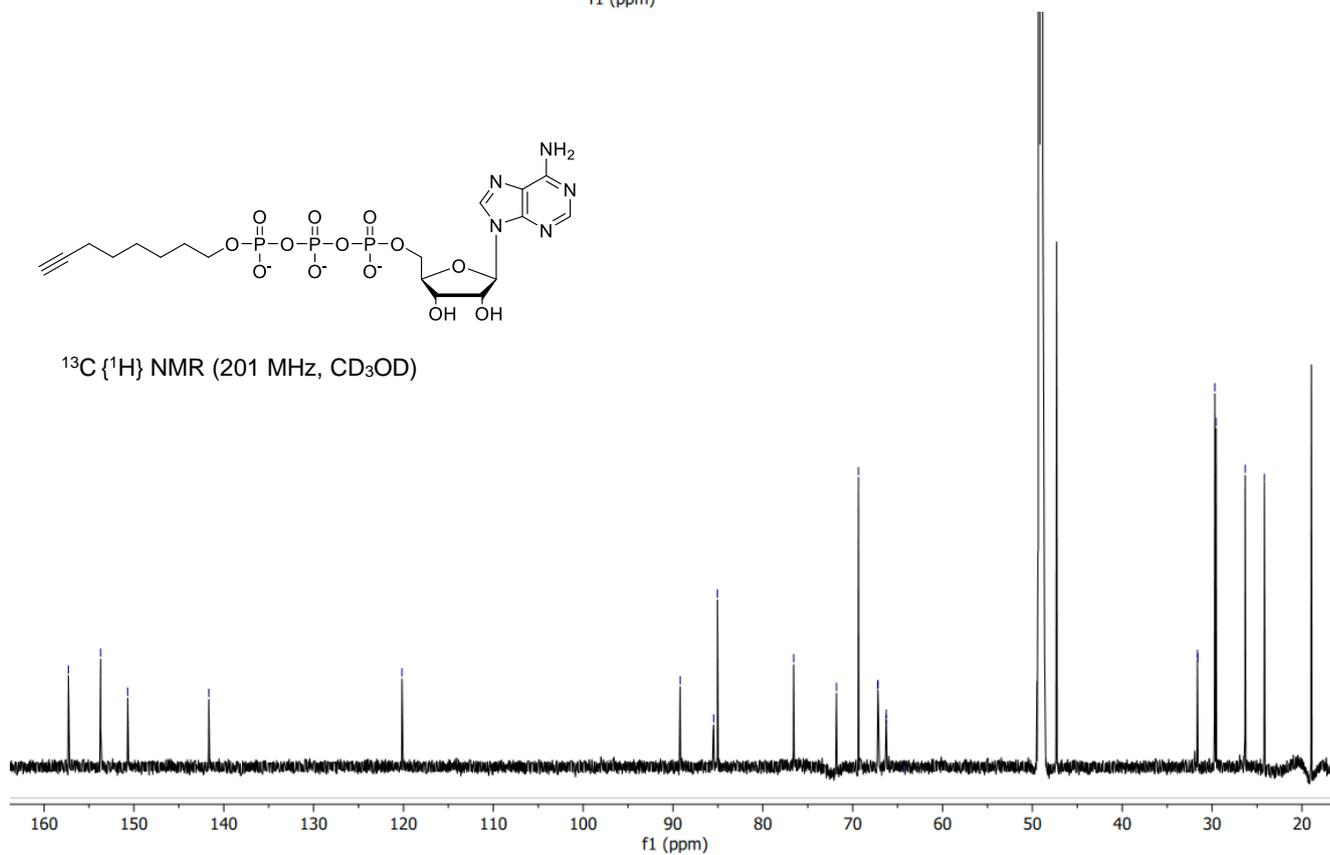
## $^1\text{H}$ , $^{13}\text{C}$ $\{^1\text{H}\}$ and $^{31}\text{P}$ $\{^1\text{H}\}$ spectra of compound 6



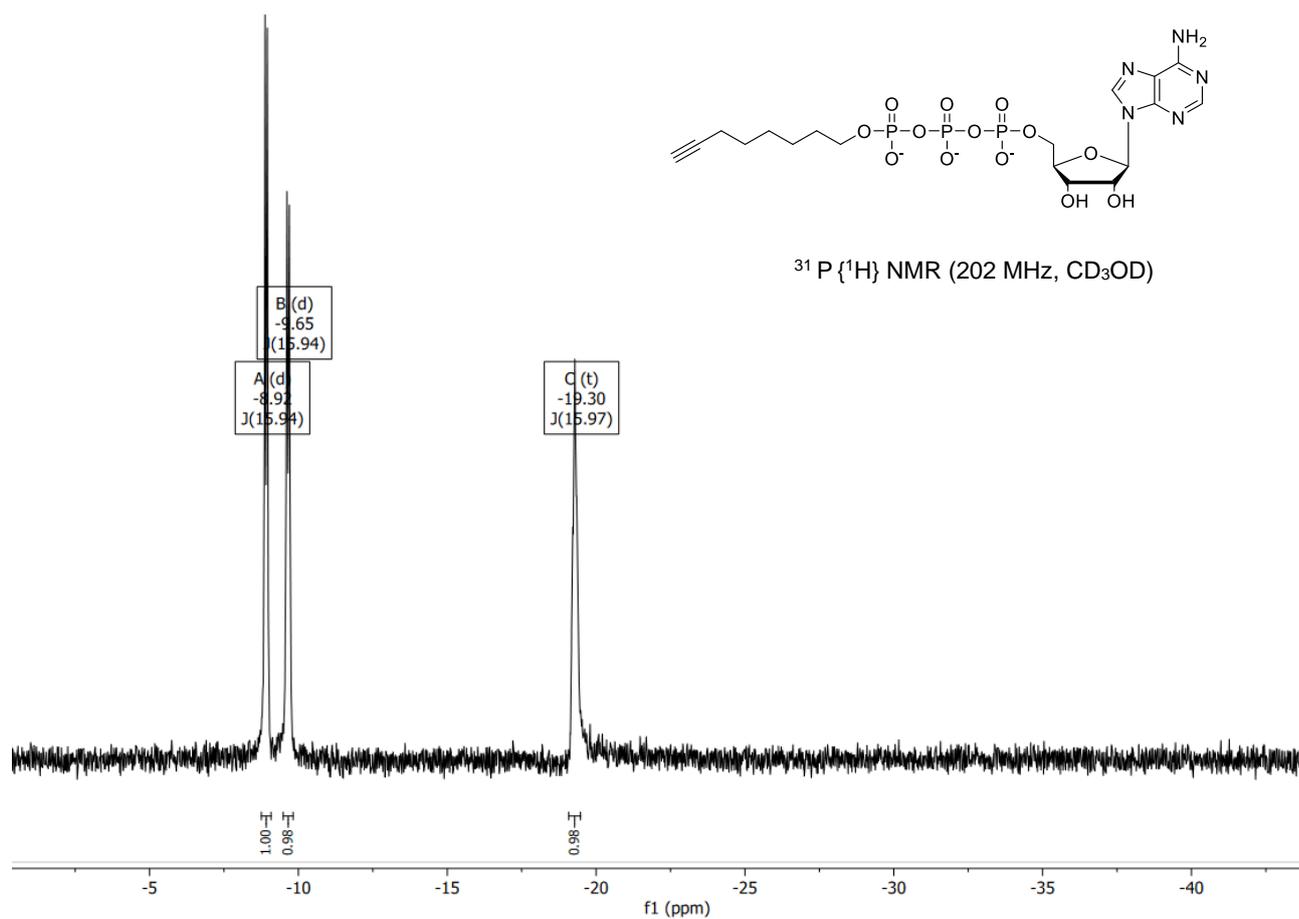
$^1\text{H}$  NMR (500 MHz,  $\text{CD}_3\text{OD}$ )



$^{13}\text{C}$   $\{^1\text{H}\}$  NMR (201 MHz,  $\text{CD}_3\text{OD}$ )

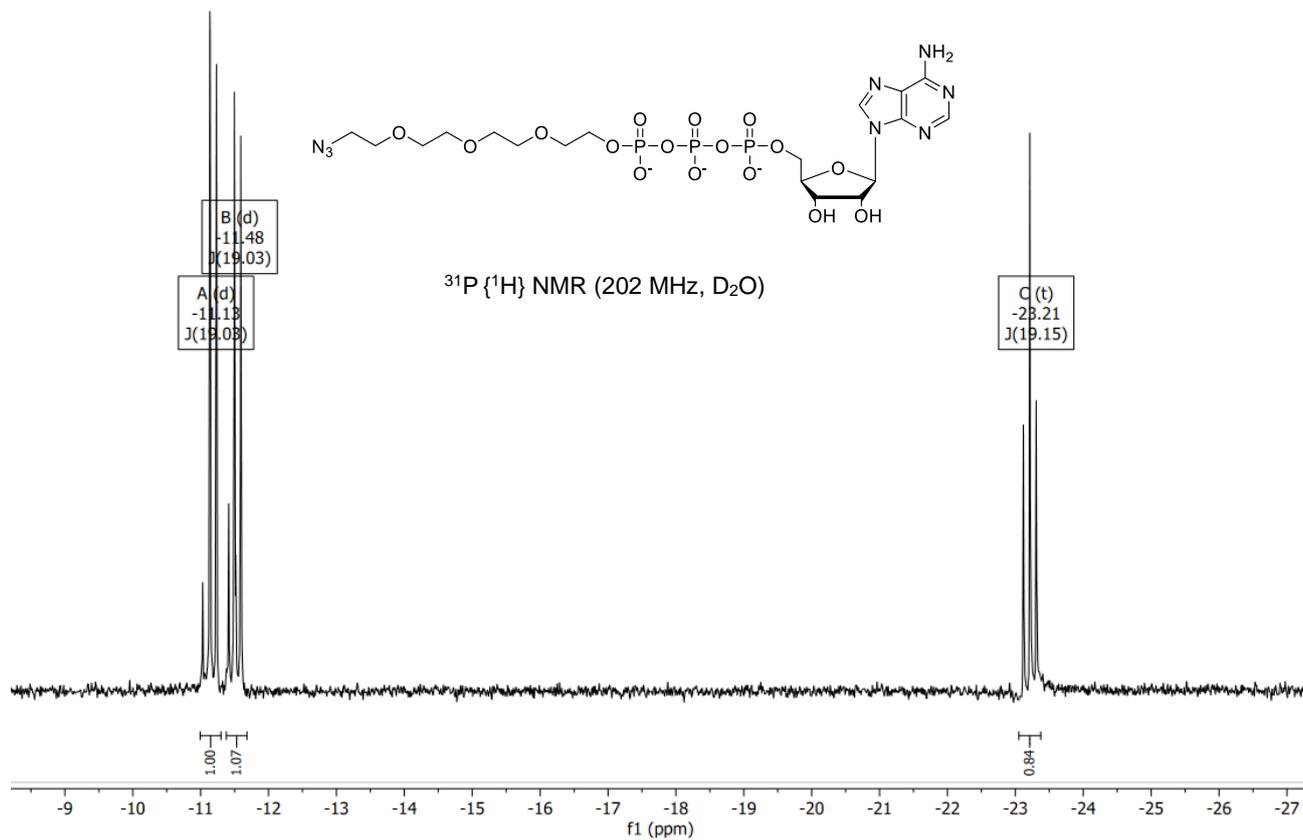
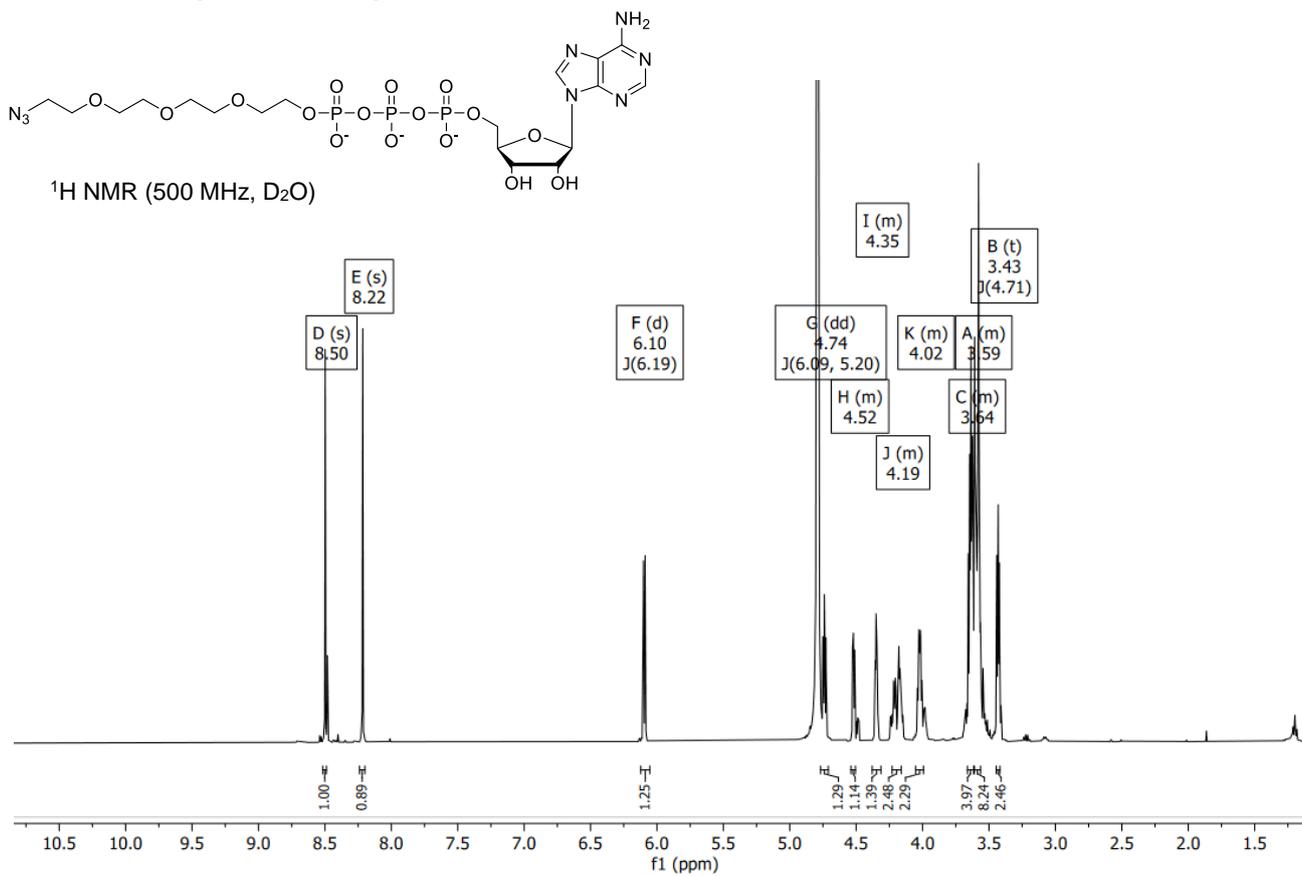


# SUPPORTING INFORMATION



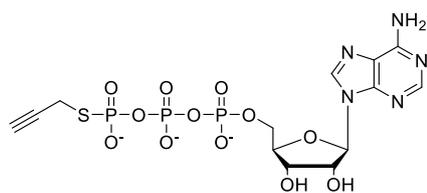
# SUPPORTING INFORMATION

## $^1\text{H}$ and $^{31}\text{P}$ $\{^1\text{H}\}$ spectra of compound 7

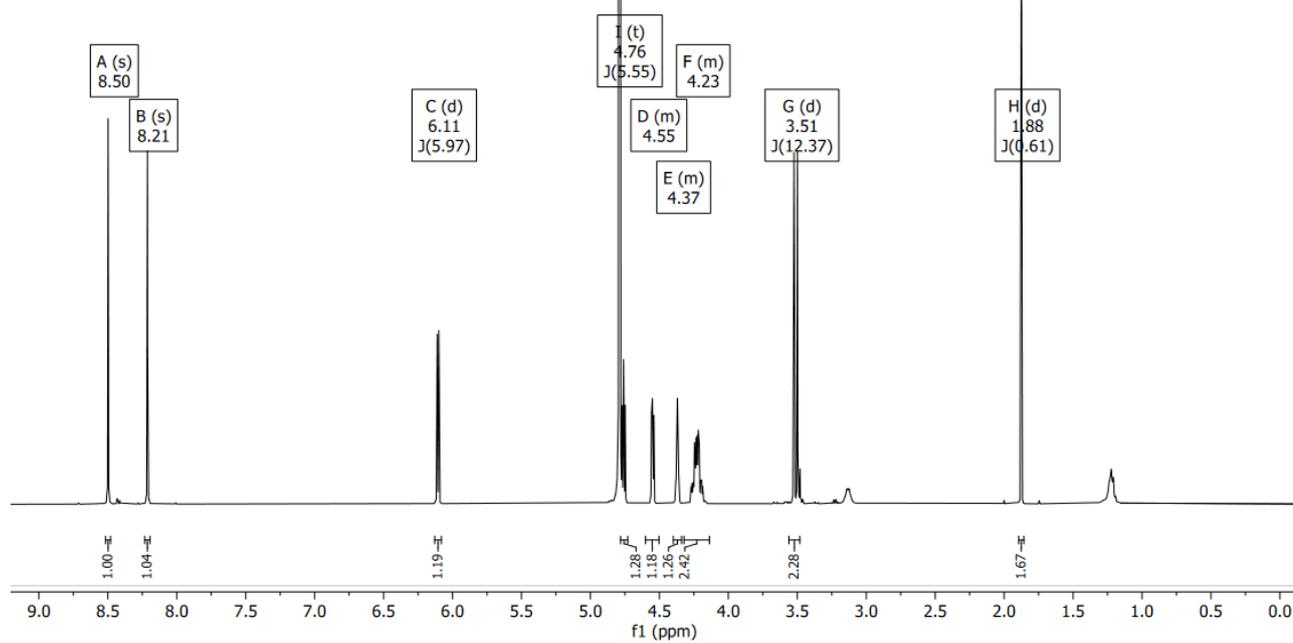


## SUPPORTING INFORMATION

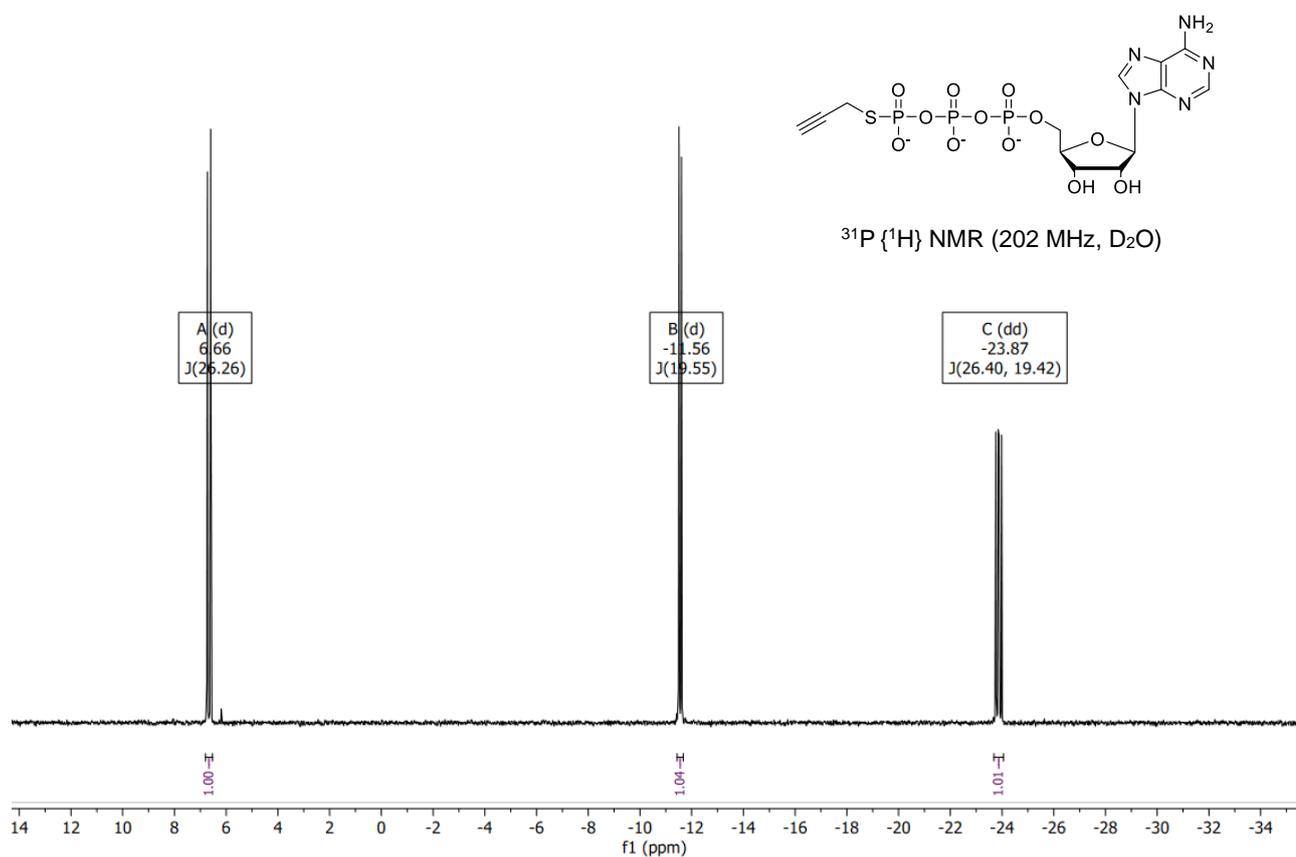
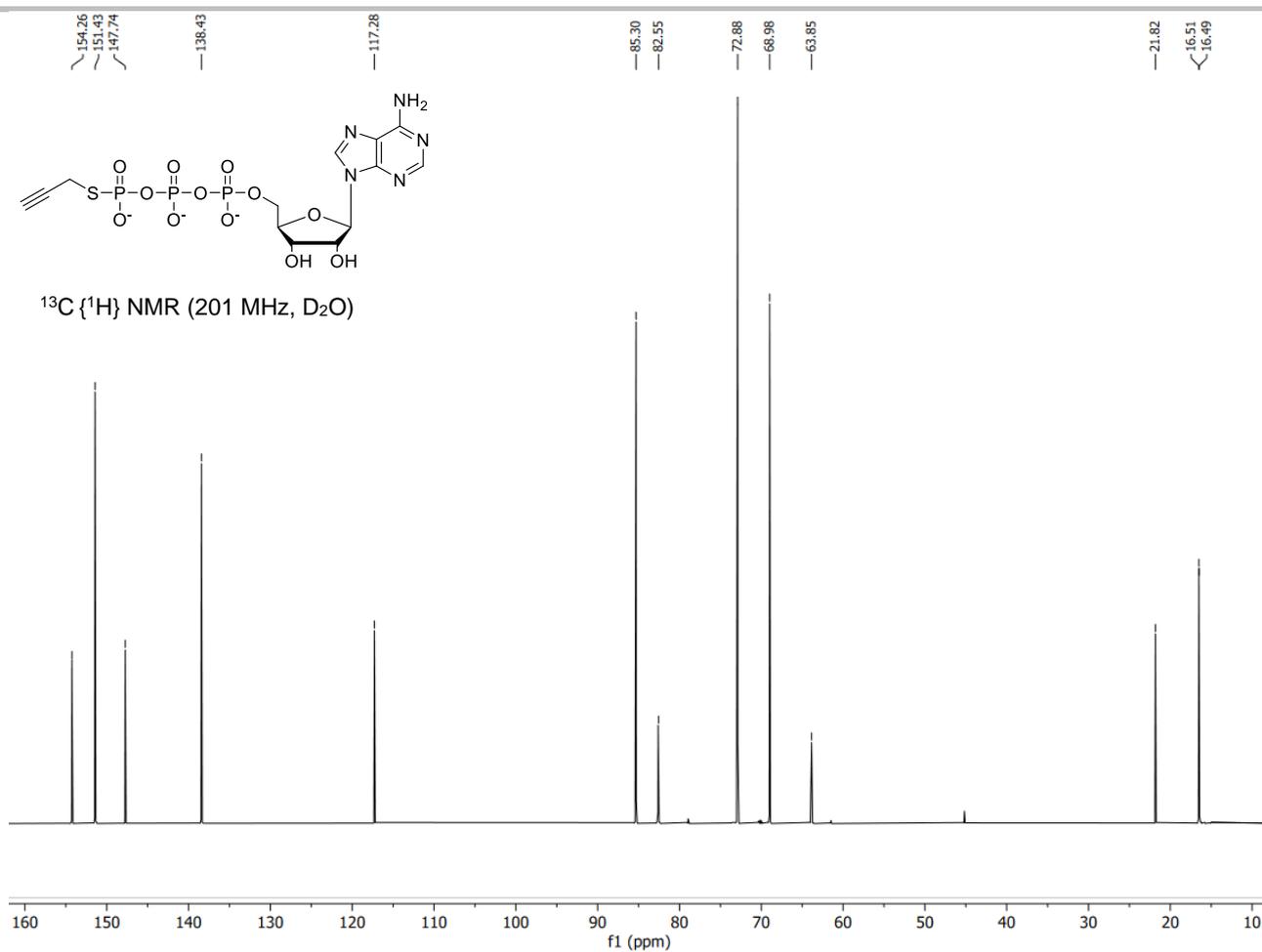
### $^1\text{H}$ , $^{13}\text{C}$ $\{^1\text{H}\}$ and $^{31}\text{P}$ $\{^1\text{H}\}$ spectra of compound 8



$^1\text{H}$  NMR (500 MHz,  $\text{D}_2\text{O}$ )



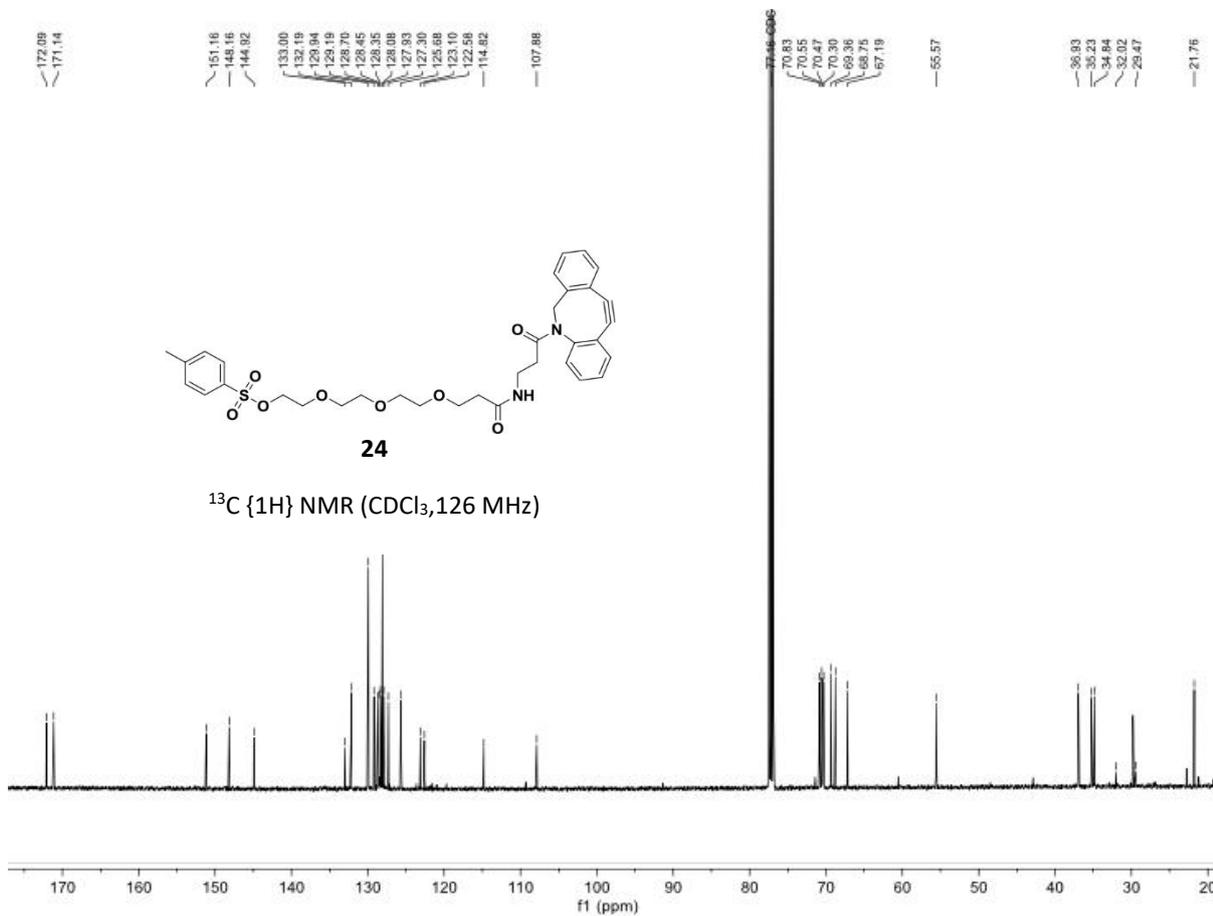
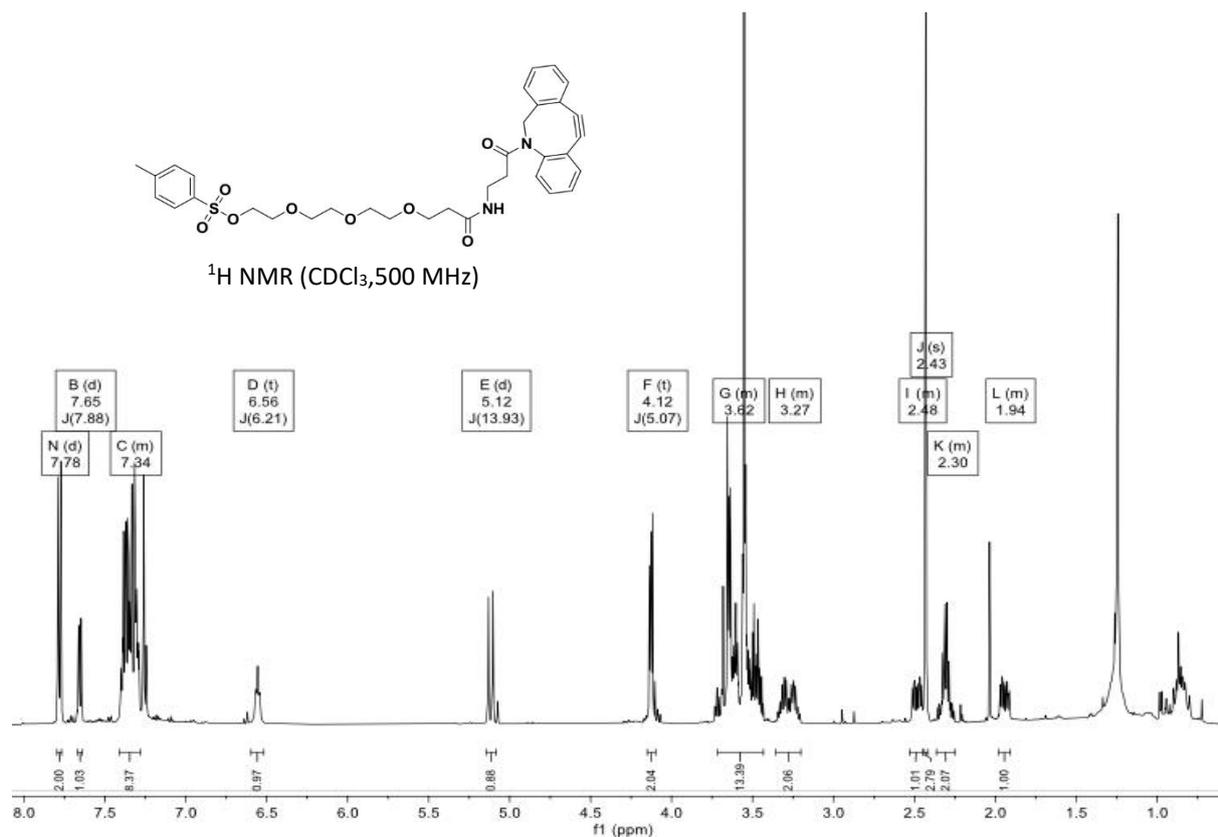
# SUPPORTING INFORMATION

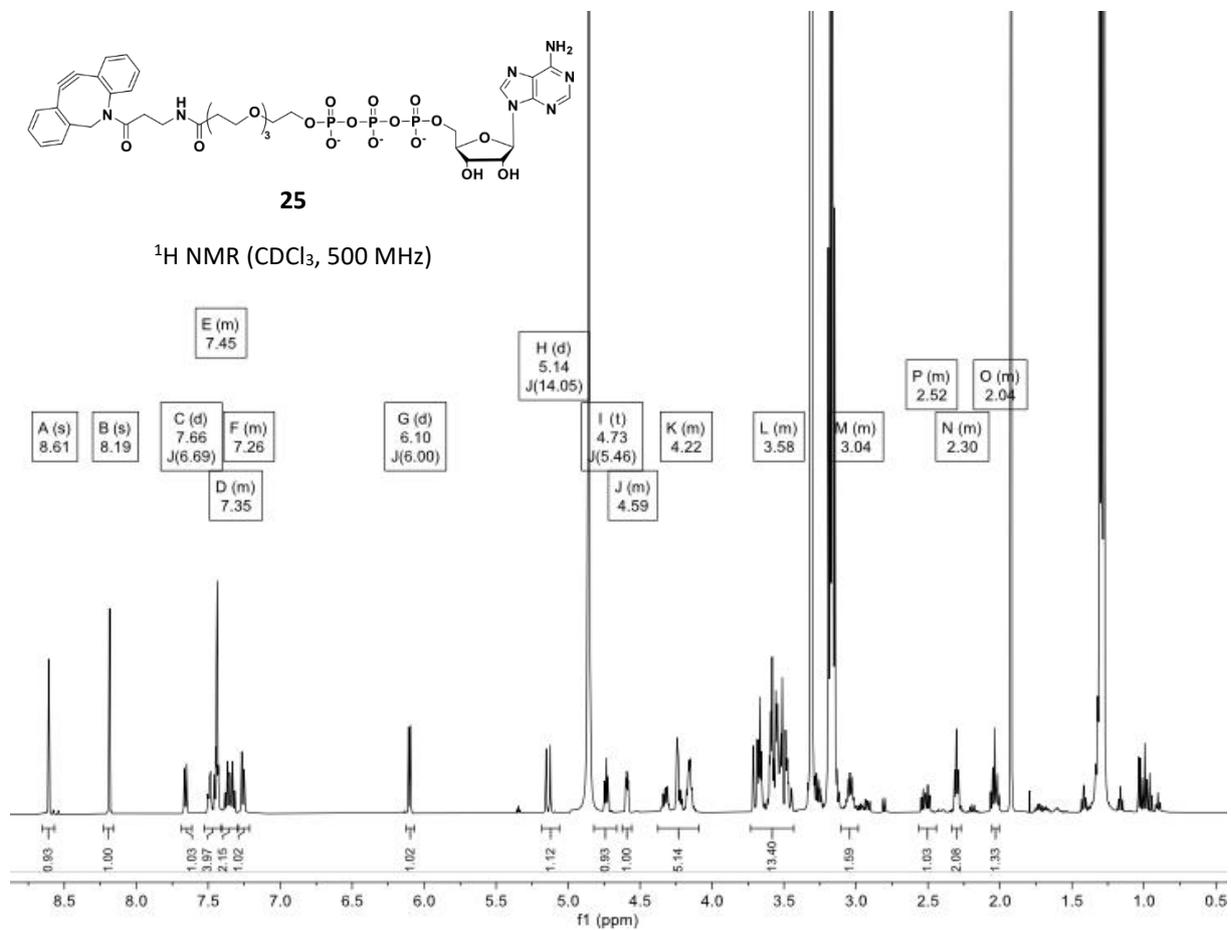
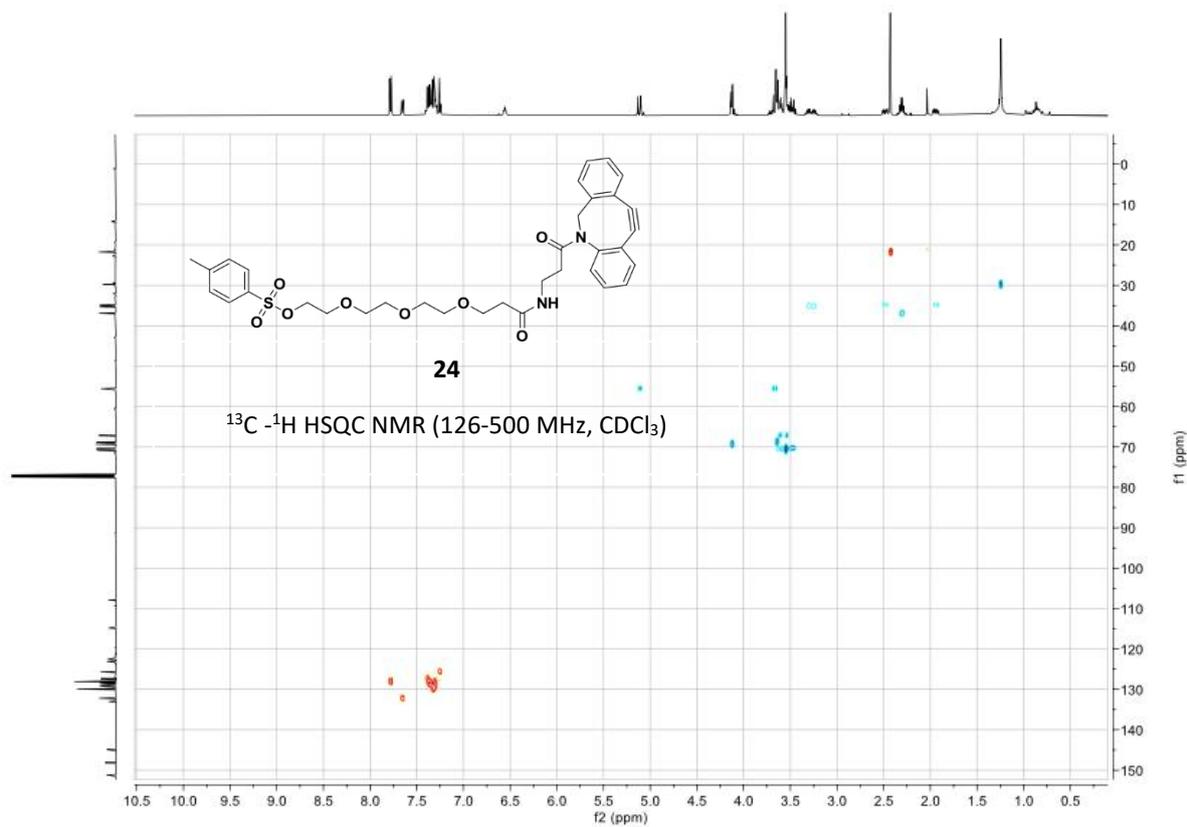


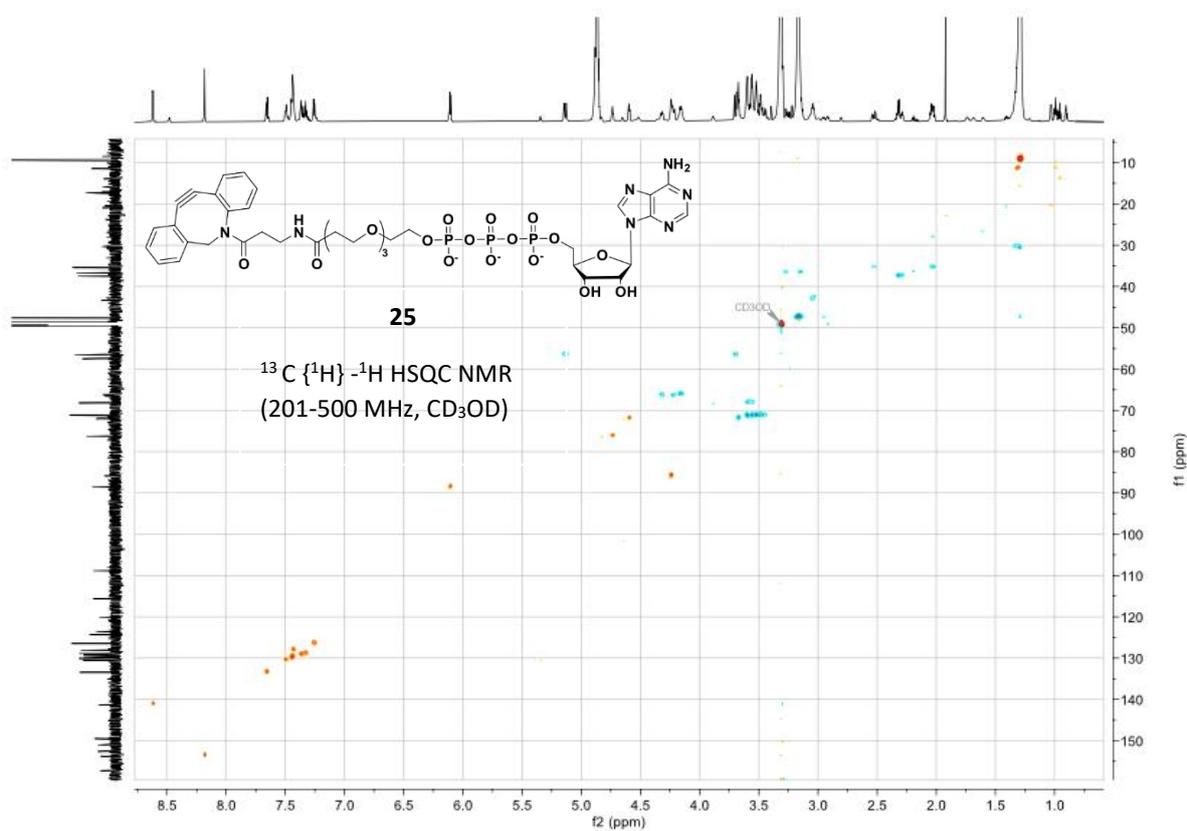
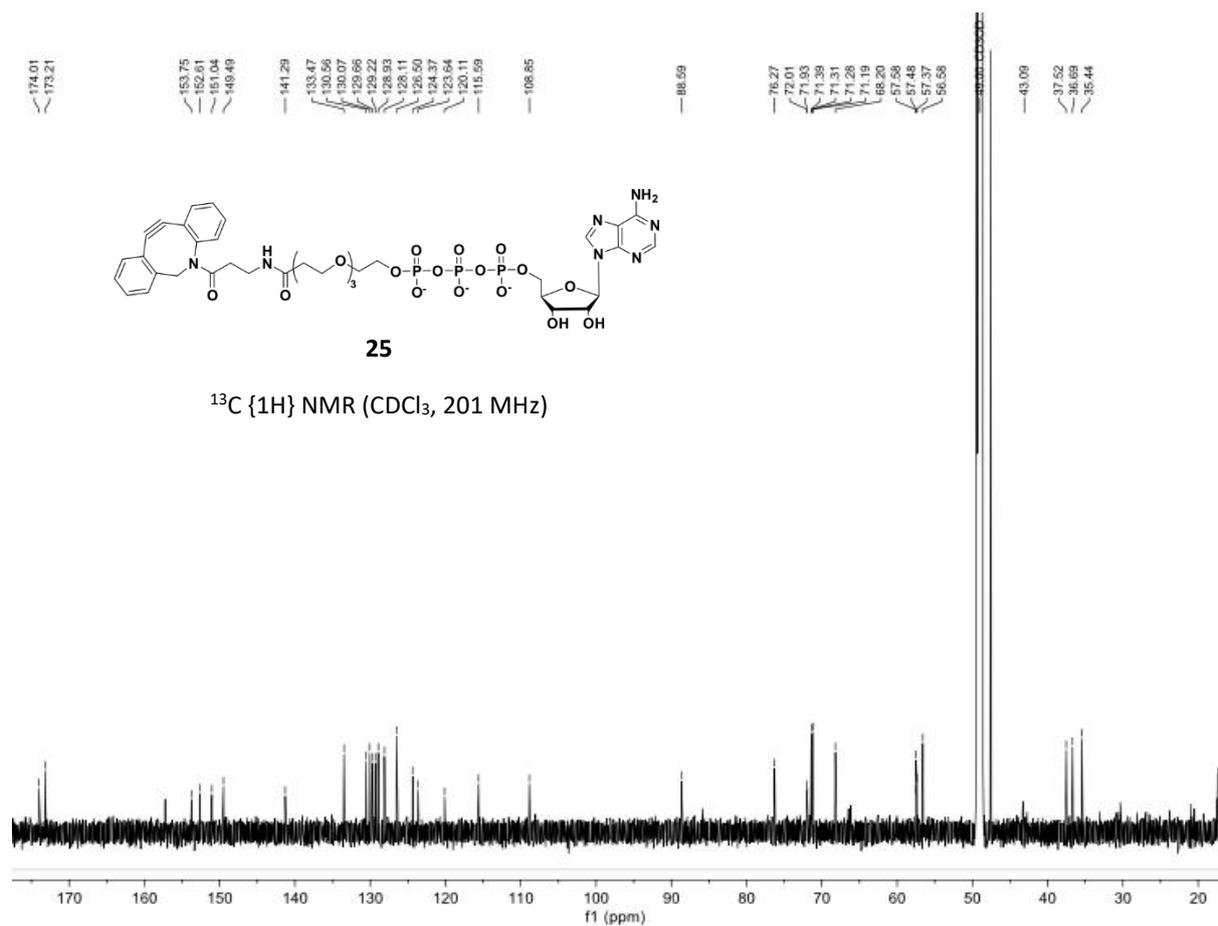
## 7. References

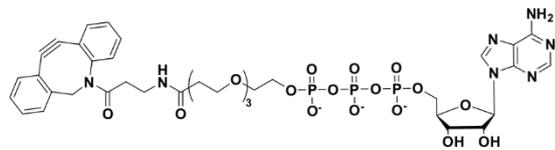
- [1] S. M. Hacker, M. Mex, A. Marx, *J. Org. Chem.* **2012**, *77*, 10450-10454.
- [2] S. M. Hacker, M. Welter, A. Marx, *Curr. Protoc. Nucleic Acid Chem.* **2015**, *60*, 13.14.11-13.14.25.
- [3] Z. Tang, K. Wang, W. Tan, C. Ma, J. Li, L. Liu, Q. Guo, X. Meng, *Nucleic Acids Res.* **2005**, *33*, e97.
- [4] J. R. Lillehaug, K. Kleppe, *Biochemistry* **1975**, *14*, 1225-1229.
- [5] S. Serdjukow, F. Kink, B. Steigenberger, M. Tomás-Gamasa, T. Carell, *Chem. Comm.* **2014**, *50*, 1861-1863.
- [6] S. E. Lee, L. M. Elphick, A. A. Anderson, L. Bonnac, E. S. Child, D. J. Mann, V. Gouverneur, *Bioorganic Med. Chem. Lett.* **2009**, *19*, 3804-3807.
- [7] A. Espinasse, X. Wen, J. D. Goodpaster, E. E. Carlson, *ACS Chem. Biol.* **2020**, *15*, 1252-1260.
- [8] L. Taemaitree, A. Shivalingam, A. H. El-Sagheer, T. Brown, *Nat. Commun.* **2019**, *10*, 1610.
- [9] K. D. Green, M. K. H. Pflum, *ChemBioChem* **2009**, *10*, 234-237.
- [10] Madhusudan, E. A. Trafny, N.-H. Xuong, J. A. Adams, L. F. T. Eyck, S. S. Taylor, J. M. Sowadski, *Protein Science* **1994**, *3*, 176-187.
- [11] T. M. Anthony, M. K. H. Pflum, *Bioorganic Med. Chem.* **2018**, *26*, 2331-2336.
- [12] J. H. Eastberg, J. Pelletier, B. L. Stoddard, *Nucleic Acids Res.* **2004**, *32*, 653-660.
- [13] U. Das, L. K. Wang, P. Smith, A. Jacewicz, S. Shuman, *Nucleic Acids Res.* **2013**, *42*, 1152-1161.
- [14] O. Trott, A. J. Olson, *J. Comput. Chem.* **2010**, *31*, 455-461.
- [15] J. Eberhardt, D. Santos-Martins, A. F. Tillack, S. Forli, *J. Chem. Inf. Model.* **2021**, *61*, 3891-3898.
- [16] R. Quiroga, M. A. Villarreal, *PLOS ONE* **2016**, *11*, e0155183.
- [17] C. C. Richardson, in *The Enzymes*, Vol. 14 (Ed.: P. D. Boyer), Academic Press, **1981**, pp. 299-314.

## 8.2 NMR Spectra of Unpublished Work



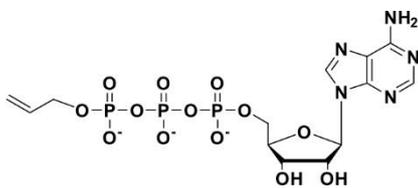
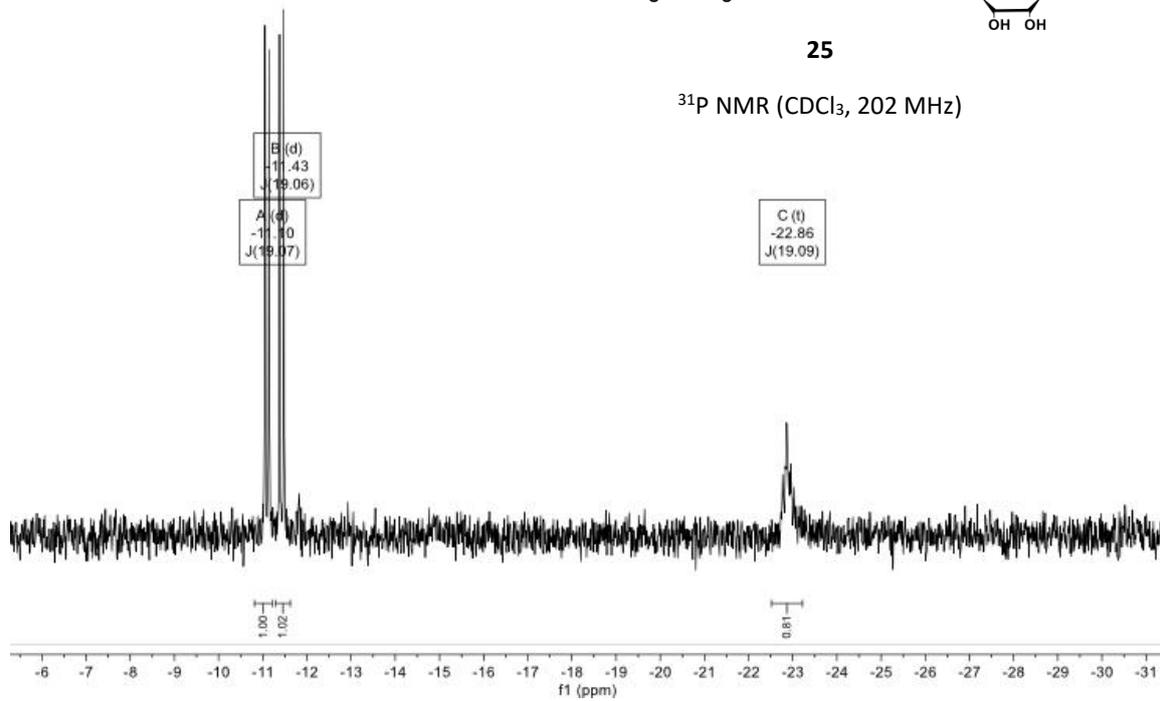






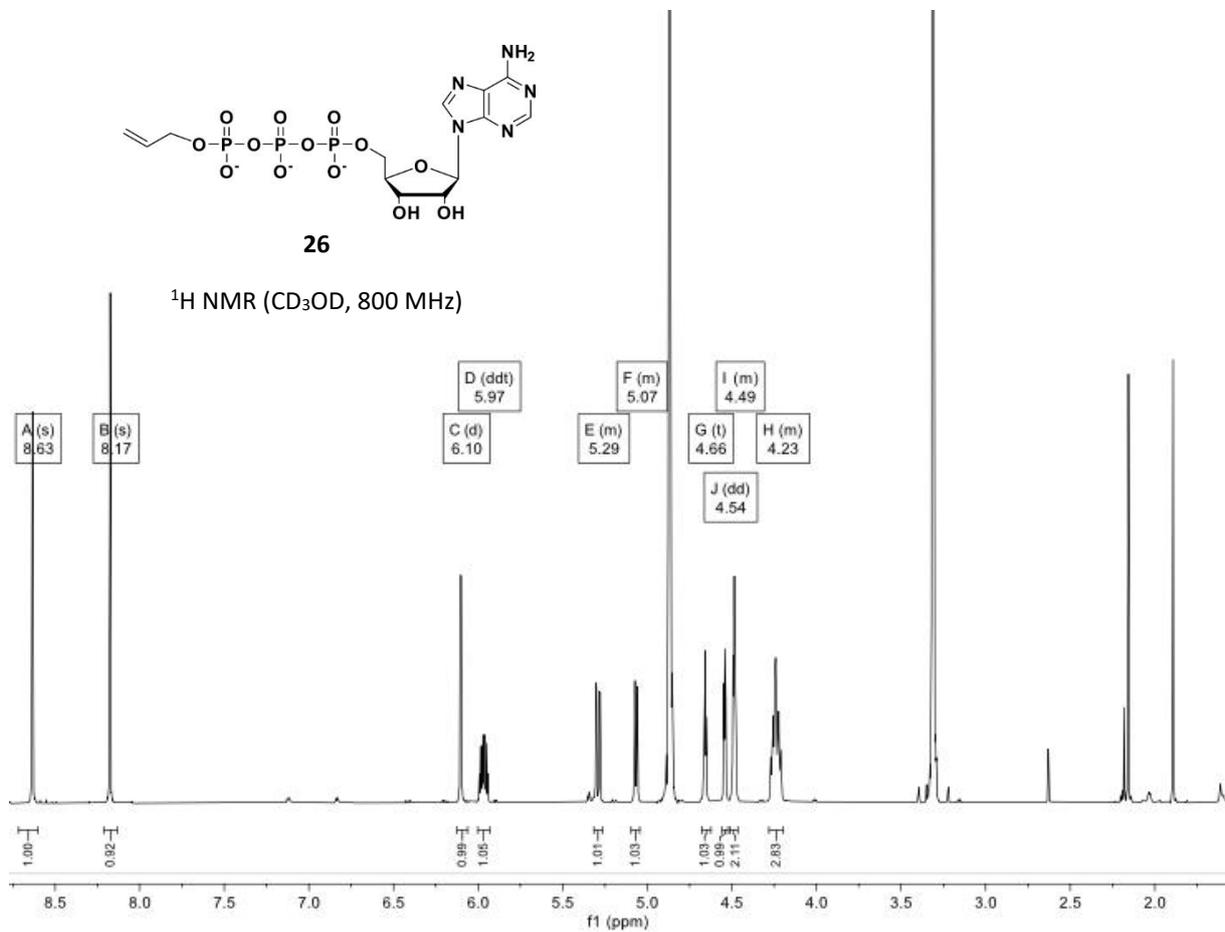
25

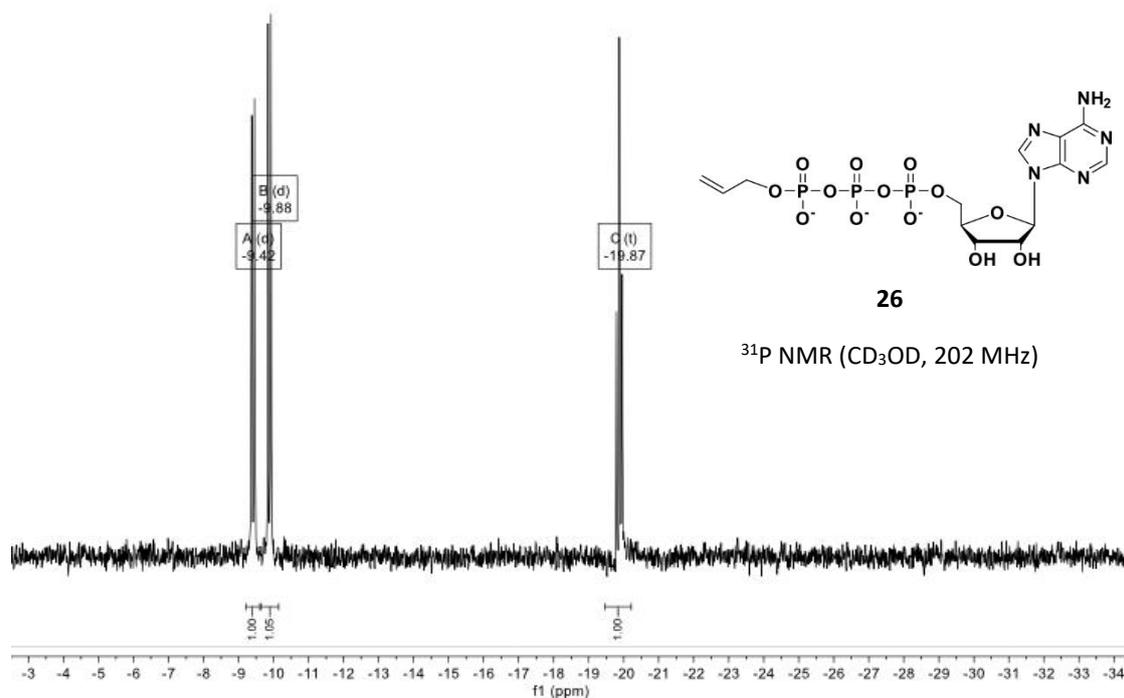
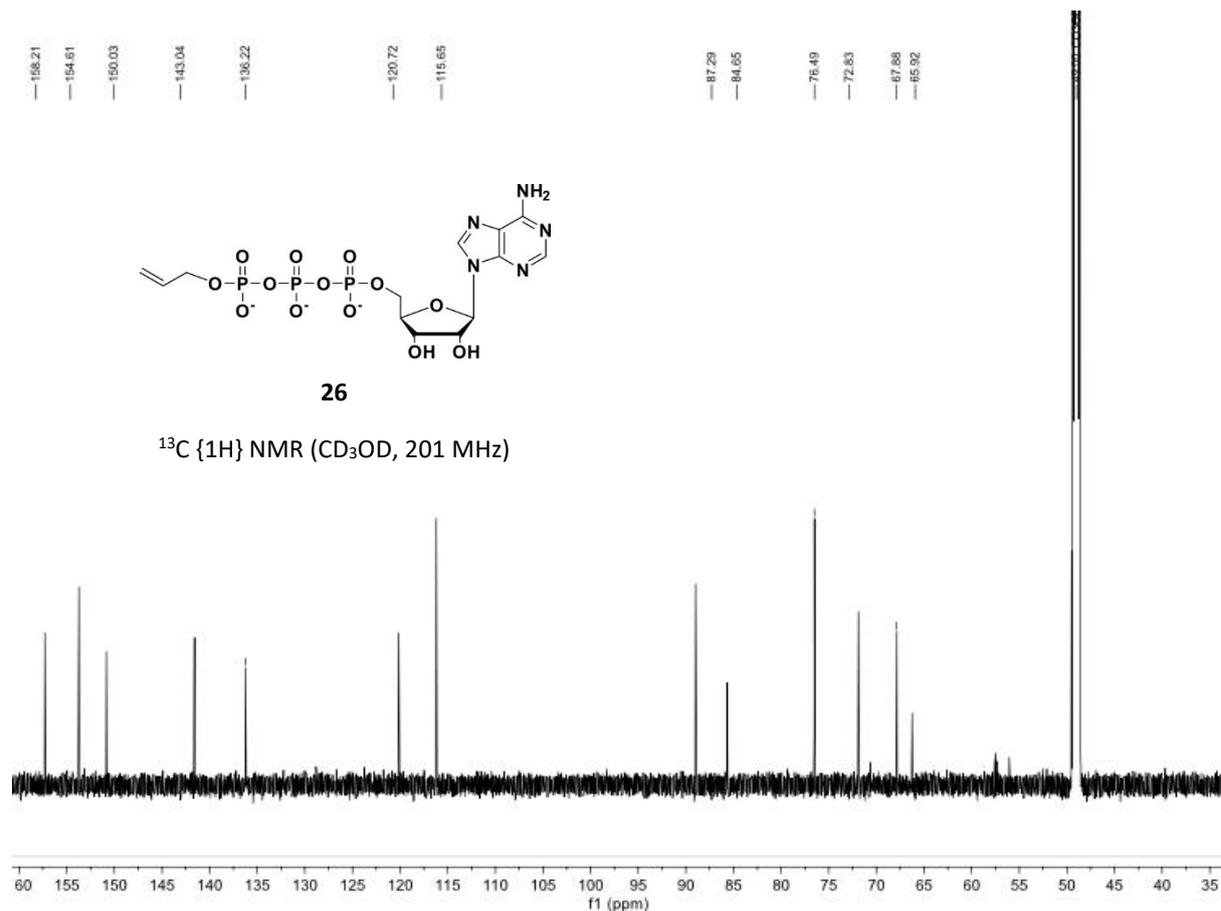
$^{31}\text{P}$  NMR ( $\text{CDCl}_3$ , 202 MHz)

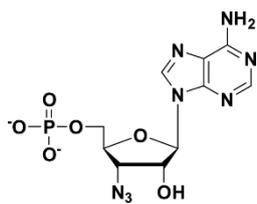


26

$^1\text{H}$  NMR ( $\text{CD}_3\text{OD}$ , 800 MHz)

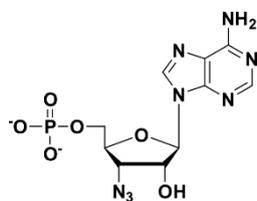
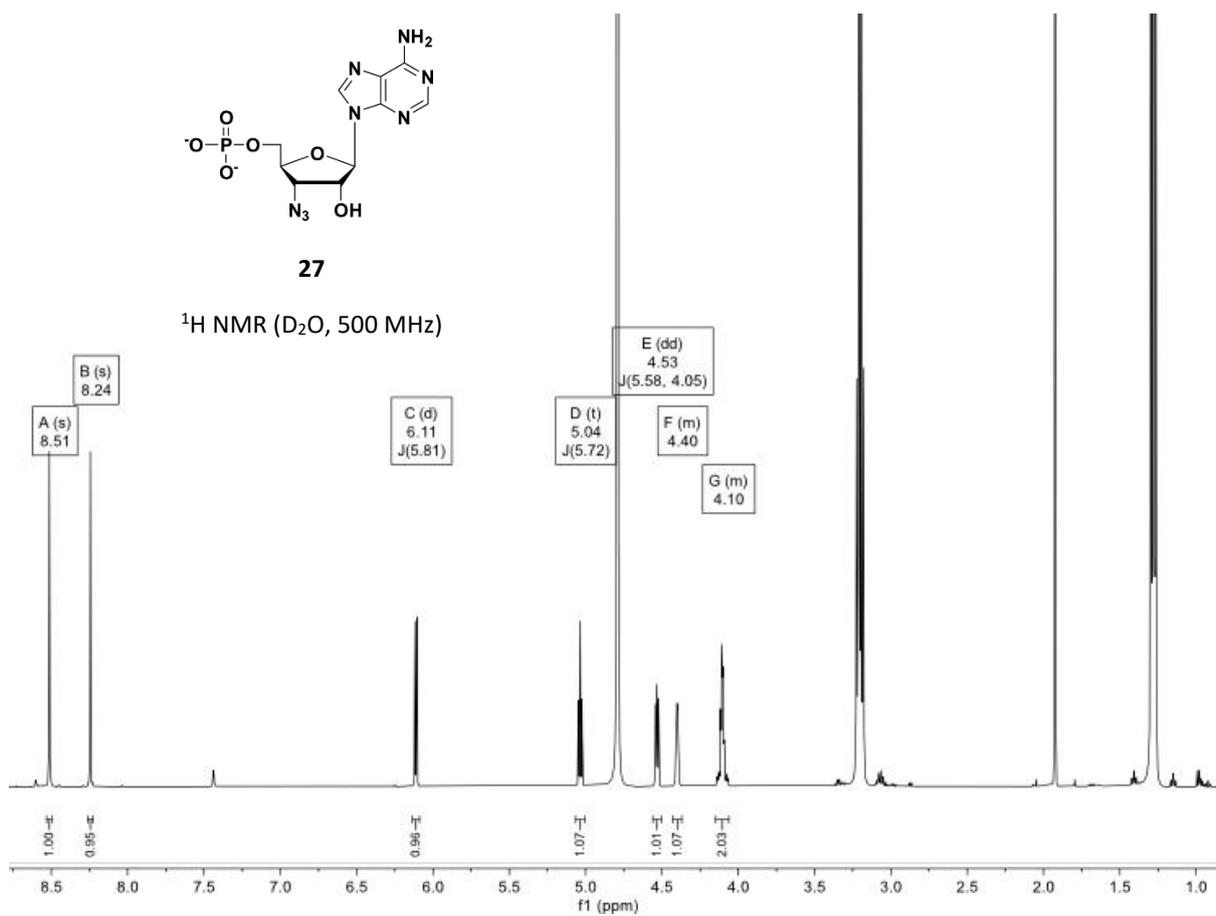






27

<sup>1</sup>H NMR (D<sub>2</sub>O, 500 MHz)



27

<sup>31</sup>P NMR (D<sub>2</sub>O, 202 MHz)

